

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C12N 15/12, C07K 14/47</b>	<b>A2</b>	<b>(11) International Publication Number:</b> <b>WO 98/11217</b> <b>(43) International Publication Date:</b> 19 March 1998 (19.03.98)
<b>(21) International Application Number:</b> PCT/JP97/03239 <b>(22) International Filing Date:</b> 12 September 1997 (12.09.97)  <b>(30) Priority Data:</b> 8/243060 13 September 1996 (13.09.96) JP  <b>(71) Applicants (for all designated States except US):</b> SAGAMI CHEMICAL RESEARCH CENTER [JP/JP]; 4-1, Nishi- Ohnuma 4-chome, Sagamihara-shi, Kanagawa 229 (JP). PROTEGENE INC. [JP/JP]; 2-20-3, Naka-cho, Meguro-ku, Tokyo 153 (JP).  <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> KATO, Seishi [JP/JP]; 3-46-50, Wakamatsu, Sagamihara-shi, Kanagawa 229 (JP). SEKINE, Shingo [JP/JP]; 4-4-1, Nishi-Ohnuma, Sagamihara-shi, Kanagawa 229 (JP). YAMAGUCHI, Tomoko [JP/JP]; 5-13-11, Takasago, Katsushika-ku, Tokyo 125 (JP). KOBAYASHI, Midori [JP/JP]; 647-2, Chougo, Fujisawa-shi, Kanagawa 252 (JP).  <b>(74) Agents:</b> AOYAMA, Tamotsu et al.; Aoyama & Partners, IMP Building, 3-7, Shiromi 1-chome, Chuo-ku, Osaka-shi, Osaka 540 (JP).		<b>(81) Designated States:</b> AU, CA, JP, MX, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>
<b>(54) Title:</b> HUMAN PROTEINS HAVING SECRETORY SIGNAL SEQUENCES AND DNAs ENCODING THESE PROTEINS		
<b>(57) Abstract</b>  [Problems to be solved] To provide human proteins having secretory signal sequences and cDNAs encoding said proteins. [Means to solve the problems] Proteins containing any of the amino acid sequences represented by Sequence No. 1 to Sequence No. 9 and DNAs encoding said proteins exemplified by cDNAs containing any of the base sequences represented by Sequence No. 10 to Sequence No. 18. Said proteins can be provided by expressing cDNAs encoding human proteins having secretory signal sequences with verified secretory functions and recombinants of these human cDNAs.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## DESCRIPTION

Human Proteins Having Secretory  
Signal Sequences and DNAs Encoding These Proteins

## TECHNICAL FIELD

The present invention relates to human proteins having secretory signal sequences and DNAs encoding these proteins. The proteins of the present invention can be used as pharmaceuticals or as antigens for preparing antibodies against said proteins. The cDNAs of the present invention can be used as probes for the gene diagnosis and gene sources for the gene therapy. Furthermore, the cDNAs can be used as gene sources for large-scale production of the proteins encoded by said cDNAs.

## BACKGROUND ART

Cells secrete many proteins outside the cells. These secretory proteins play important roles for the proliferation control, the differentiation induction, the material transportation, the biological protection, etc. in the cells. Different from intracellular proteins, the secretory proteins exert their actions outside the cells, whereby they can be administered in the intracorporeal manner such as the injection or the drip to anticipate the potentialities as medicines. In fact, a number of human secretory proteins such as interleukins, interferons, erythropoietin, thrombolytic agents, etc. have been currently utilized as



medicines. In addition, secretory proteins other than those described above have been undergoing clinical trials to develop as pharmaceuticals. Since it has been conceived that the human cells still produce many unknown secretory proteins, availability of these secretory proteins as well as genes encoding them is expected to lead to the development of novel pharmaceuticals using these proteins.

Heretofore, such a secretory protein has been obtained by a method comprising the isolation and purification of the target protein from a large amount of the blood or a cell culture supernatant by using the biological activity as an indicator, determination of its primary structure followed by cloning of the corresponding cDNA on the basis of the information on the thus-obtained amino acid sequence, and production of the recombinant protein using said cDNA. However, the contents of the secretory proteins are generally so low that the isolation and purification are difficult in many cases. On the other hand, secretory proteins and type-I membrane proteins possess hydrophobic sequences, defined as the secretory signal sequences, consisting of about 20 amino acid residues at the amino acid termini (the N-termini). Therefore, the cloning of genes encoding the secretory proteins or type-I membrane proteins is expected to be performed by using the presence or the absence of these secretory signal sequences as indicators.

#### DISCLOSURE OF INVENTION

The object of the present invention is to provide novel human proteins having secretory signal sequences and DNAs

encoding said proteins.

As the result of intensive studies, the present inventors were successful in cloning of cDNAs having secretory signal sequences from a human full-length cDNA bank, thereby completing the present invention. That is to say, the present invention provides proteins containing any of the amino acid sequences represented by Sequence No. 1 to Sequence No. 9 that are human proteins having secretory signal sequences. The present invention, also, provides DNAs encoding said proteins exemplified as cDNAs containing any of the base sequences represented by Sequence No. 10 to sequence No. 18.

Each of the proteins of the present invention can be obtained, for example, by a method for isolation from human organs, cell lines, etc, a method for preparation of the peptide by the chemical synthesis on the basis of the amino acid sequence of the present invention, or a method for production with the recombinant DNA technology using the DNA encoding the human secretory protein of the present invention, wherein the method for obtainment by the recombinant DNA technology is employed preferably. For example, an in vitro expression can be achieved by preparation of an RNA by the in vitro transcription from a vector having a cDNA of the present invention, followed by the in vitro translation using this RNA as a template. Also, the recombination of the translation domain to a suitable expression vector by the method known in the art leads to the expression of a large amount of the encoded protein by using *Escherichia coli*, *Bacillus subtilis*, yeasts, animal cells, and so on.

In the case in which a protein of the present invention is expressed by a microorganism such as *Escherichia coli*, the translation region of a cDNA of the present invention is constructed in an expression vector having an origin, a promoter, ribosome-binding site(s), cDNA-cloning site(s), a terminator, etc. that can be replicated in the microorganism and, after transformation of the host cells with said expression vector, the thus-obtained transformant is incubated, whereby the protein encoded by said cDNA can be produced on a large scale in the microorganism. In that case, a maturation protein can be obtained by performing the expression with inserting an initiation codon in the translation region where the secretory signal sequence is removed. Alternatively, a fusion protein with another protein can be expressed. Only a protein portion encoding said cDNA can be obtained by cleavage of said fusion protein with an appropriate protease.

In the case in which a protein of the present invention is secretory-expressed in animal cells, the protein of the present invention can be secretory-produced as a maturation protein outside the cells, when the translation region of said cDNA is subjected to recombination to an expression vector for animal cells that has a promoter for the animal cells, a splicing domain, a poly(A) addition site, etc., followed by transfection into the animal cells.

The proteins of the present invention include peptide fragments (more than 5 amino acid residues) containing any partial amino acid sequence of the amino acid sequences represented by Sequence No. 1 to Sequence No. 9. These

fragments can be used as antigens for preparation of the antibodies. Also, the proteins of the present invention are secreted in the form of maturation proteins outside the cells, after the signal sequences are removed. Therefore, these maturation proteins shall come within the scope of the present invention. The N-terminal amino acid sequences of the maturation proteins can be easily identified by using the method for the cleavage-site determination in a signal sequence [Japanese Patent Kokai Publication No. 1996-187100]. Furthermore, many secretory proteins are subjected to the processing after the secretion to be converted to the active forms. These activated proteins or peptides shall come within the scope of the present invention. When glycosylation sites are present in the amino acid sequences, expression in appropriate animal cells affords glycosylated proteins. Therefore, these glycosylated proteins or peptides also shall come within the scope of the present invention.

The DNAs of the present invention include all DNAs encoding the above-mentioned proteins. Said DNAs can be obtained using the method by chemical synthesis, the method by cDNA cloning, and so on.

Each of the cDNAs of the present invention can be cloned from, for example, a cDNA library of the human cell origin. The cDNA is synthesized using as a template a poly(A)<sup>+</sup> RNA extracted from human cells. The human cells may be cells delivered from the human body, for example, by the operation or may be the culture cells. The cDNA can be synthesized by using any method selected from the Okayama-Berg method [Okayama, H. and Berg, P., Mol. Cell. Biol. 2: 161-170

(1982)], the Gubler-Hoffman method [Gubler, U. and Hoffman, J. Gene 25: 263-269 (1983)], and so on, but it is preferred to use the capping method [Kato, S. et al., Gene 150: 243-250 (1994)] as illustrated in Examples in order to obtain a full-length clone in an effective manner.

The primary selection of a cDNA encoding a human protein having a secretory signal sequence is performed by the sequencing of a partial base sequence of the cDNA clone selected at random from the cDNA library, sequencing of the amino acid sequence encoded by the base sequence, and recognition of the presence or absence of hydrophobic site(s) in the resulting N-terminal amino acid sequence region. Next, the secondary selection is carried out by determination of the whole base sequence by the sequencing and the protein expression by the in vitro translation. The ascertainment of the cDNA of the present invention for encoding the protein having the secretory signal sequence is performed by using the signal sequence detection method [Yokoyama-Kobayashi, M. et al., Gene 163: 193-196 (1995)]. In other words, the ascertainment for the coding portion of the inserted cDNA fragment to function as a signal sequence is provided by fusing a cDNA fragment encoding the N-terminus of the target protein with a cDNA encoding the protease domain of urokinase and then expressing the resulting cDNA in COS7 cells to detect the urokinase activity in the cell culture medium.

The cDNAs of the present invention are characterized by containing any of the base sequences represented by Sequence No. 10 to Sequence No. 18 or any of the base sequences represented by Sequence No. 19 to Sequence No. 27. Table 1

summarizes the clone number (HP number), the cells affording the cDNA, the total base number of the cDNA, and the number of the amino acid residues of the encoded protein, for each of the cDNAs.

Table 1

Sequence Number	HP Number	Cells	Number of Bases	Number of Amino Acid Residues
1. 10. 19	HP00658	HT-1080	1296	154
2. 11. 20	HP00714	KB	3311	315
3. 12. 21	HP00876	Stomach cancer	1152	158
4. 13. 22	HP01134	Liver	1749	376
5. 14. 23	HP10029	KB	988	173
6. 15. 24	HP10189	KB	390	93
7. 16. 25	HP10269	U937	4667	1172
8. 17. 26	HP10298	Stomach cancer	1086	122
9. 18. 27	HP10368	Stomach cancer	866	175

Hereupon, the same clone as any of the cDNAs of the present invention can be easily obtained by screening of the cDNA library constructed from the cell line or the human tissue employed in the present invention, by the use of an oligonucleotide probe synthesized on the basis of the corresponding cDNA base sequence depicted in Sequence No. 19 to Sequence No. 27.

In general, the polymorphism due to the individual difference is frequently observed in human genes. Therefore, any cDNA that is subjected to insertion or deletion of one or

plural nucleotides and/or substitution with other nucleotides in Sequence No. 10 to Sequence No. 27 shall come within the scope of the present invention.

In a similar manner, any protein that is produced by these modifications comprising insertion or deletion of one or plural nucleotides and/or substitution with other nucleotides shall come within the scope of the present invention, as far as said protein possesses the activity of the corresponding protein having the amino acid sequence represented by Sequence No. 1 to Sequence No. 9.

The cDNAs of the present invention include cDNA fragments (more than 10 bp) containing any partial base sequence of the base sequence represented by Sequence No. 10 to No. 18 or of the base sequence represented by Sequence No. 19 to No. 27. For example, as illustrated in Examples, the portion encoding the secretory signal sequence can be employed as means to secrete an optionally selected protein outside the cells by fusing with a cDNA encoding another protein. Also, DNA fragments consisting of a sense chain and an anti-sense chain shall come within this scope. These DNA fragments can be used as the probes for the gene diagnosis.

#### BRIEF DESCRIPTION OF DRAWINGS

Figure 1: A figure depicting the structure of the secretory signal sequence detection vector pSSD3.

Figure 2: A figure depicting the construction of the secretory signal sequence - the urokinase fusion gene.

Figure 3: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded

by clone HP00685.

Figure 4: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP00714.

Figure 5: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP00876.

Figure 6: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP01134.

Figure 7: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10029.

Figure 8: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10189.

Figure 9: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10269.

Figure 10: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10298.

Figure 11: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10368.

BEST MODE FOR CARRING OUT INVENTION

#### EXAMPLE

The present invention is embodied in more detail by the



following examples, but this embodiment is not intended to restrict the present invention. The basic operations and the enzyme reactions with regard to the DNA recombination are carried out according to the literature ["Molecular Cloning. A Laboratory Manual", Cold Spring Harbor Laboratory, 1989]. Unless otherwise stated, restrictive enzymes and a variety of modification enzymes to be used were those available from Takara Shuzo Co., Ltd. The manufacturer's instructions were used for the buffer compositions as well as for the reaction conditions, in each of the enzyme reactions. The cDNA synthesis was carried out according to the literature [Kato, S. et al., Gene 150: 243-250 (1994)].

(1) Preparation of Poly(A)<sup>+</sup> RNA

The fibrosarcoma cell line HT-1080 (ATCC CCL 121), the epidermoid carcinoma cell line KB (ATCC CRL 17), the histiocyte lymphoma cell line U937 (ATCC CRL 1593) stimulated by phorbol esters, tissues of stomach cancer delivered by the operation, and liver were used for human cells to extract mRNAs. Each of the cell lines was cultured by a conventional procedure.

After about 1 g of human tissues was homogenized in 20 ml of a 5.5 M guanidinium thiocyanate solution, total mRNAs were prepared in accordance with the literature [Okayama, H. et al., "Methods in Enzymology" Vol. 164, Academic Press, 1987]. These mRNAs were subjected to chromatography using an oligo(dT)-cellulose column washed with 20 mM Tris-hydrochloric acid buffer solution (pH 7.6), 0.5 M NaCl, and 1 mM EDTA to obtain a poly(A)<sup>+</sup> RNA in accordance with the above-mentioned literature.

## (2) Construction of cDNA Library

To a solution of 10  $\mu$ g of the above-mentioned poly(A)<sup>+</sup> RNA in 100 mM Tris-hydrochloric acid buffer solution (pH 8) was added one unit of an RNase-free, bacterium-origin alkaline phosphatase and the resulting solution was allowed to react at 37°C for one hour. After the reaction solution underwent the phenol extraction followed by the ethanol precipitation, the obtained pellets were dissolved in a mixed solution of 50 mM sodium acetate (pH 6), 1 mM EDTA, 0.1% 2-mercaptoethanol, and 0.01% Triton X-100. Thereto was added one unit of a tobacco-origin pyrophosphatase (Epicenter Technologies) and the resulting solution at a total volume of 100  $\mu$ l was allowed to react at 37°C for one hour. After the reaction solution underwent the phenol extraction followed by the ethanol precipitation, the thus-obtained pellets were dissolved in water to obtain a decapped poly(A)<sup>+</sup> RNA solution.

To a solution of the decapped poly(A)<sup>+</sup> RNA and 3 nmol of a DNA-RNA chimeric oligonucleotide (5'-dG-dG-dG-dG-dA-dA-dT-dT-dC-dG-dA-G-G-A-3') in a mixed aqueous solution of 50 mM Tris-hydrochloric acid buffer solution (pH 7.5), 0.5 mM ATP, 5 mM MgCl<sub>2</sub>, 10 mM 2-mercaptoethanol, and 25% polyethylene glycol were added 50 units of T4 RNA ligase and the resulting solution at a total volume of 30  $\mu$ l was allowed to react at 20°C for 12 hours. After the reaction solution underwent the phenol extraction followed by the ethanol precipitation, the thus-obtained pellets were dissolved in water to obtain a chimeric oligo-capped poly(A)<sup>+</sup> RNA.

After the vector pKA1 developed by the present inventors (Japanese Patent Kokai Publication No. 1992-117292) was digested with KpnI, an about 60-dT tail was inserted by a terminal transferase. This product was digested with EcoRV to remove the dT tail at one side and the resulting molecule was used as a vectorial primer.

After 6  $\mu$ g of the previously-prepared chimeric oligo-capped poly(A)<sup>+</sup> RNA was annealed with 1.2  $\mu$ g of the vectorial primer, the product was dissolved in a mixed solution of 50 mM Tris-hydrochloric acid buffer solution (pH 8.3), 75 mM KCl, 3 mM MgCl<sub>2</sub>, 10 mM dithiothreitol, and 1.25 mM dNTP (dATP + dCTP + dGTP + dTTP), mixed with 200 units of a reverse transferase (GIBCO-BRL), and the resulting solution at a total volume of 20  $\mu$ l was allowed to react at 42°C for one hour. After the reaction solution underwent the phenol extraction followed by the ethanol precipitation, the thus-obtained pellets were dissolved in a mixed solution of 50 mM Tris-hydrochloric acid buffer solution (pH 7.5), 100 mM NaCl, 10 mM MgCl<sub>2</sub>, and 1 mM dithiothreitol. Thereto were added 100 units of EcoRI and the resulting solution at a total volume of 20  $\mu$ l was allowed to react at 37°C for one hour. After the reaction solution underwent the phenol extraction followed by the ethanol precipitation, the obtained pellets were dissolved in a mixed solution of 20 mM Tris-hydrochloric acid buffer solution (pH 7.5), 100 mM KCl, 4 mM MgCl<sub>2</sub>, 10 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, and 50  $\mu$ g/ml bovine serum albumin. Thereto were added 60 units of *Escherichia coli* DNA ligase and the resulting solution was allowed to react at 16°C for 16 hours.

To the reaction solution were added 2  $\mu$ l of 2 mM dNTP, 4 units of *Escherichia coli* DNA polymerase I, and 0.1 unit of *Escherichia coli* DNase H and the resulting solution was allowed to react at 12°C for one hour and then at 22°C for one hour.

Next, the cDNA-synthesis reaction solution was used to transform *Escherichia coli* DH12S (GIBCO-BRL). The transformation was carried out by the electroporation method. A portion of the transformant was inoculated on a 2xYT agar culture medium containing 100  $\mu$ g/ml ampicillin, which was incubated at 37°C overnight. A colony grown on the culture medium was randomly picked up and inoculated on 2 ml of the 2xYT culture medium containing 100  $\mu$ g/ml ampicillin, which was incubated at 37°C overnight. The culture medium was centrifuged to separate the cells, from which a plasmid DNA was prepared by the alkaline lysis method. After the plasmid DNA was double-digested with EcoRI and NotI, the product was subjected to 0.8% agarose gel electrophoresis to determine the size of the cDNA insert. In addition, by the use of the obtained plasmid as a template, the sequence reaction using M13 universal primer labeled with a fluorescent dye and Taq polymerase (a kit of Applied Biosystems Inc.) was carried out and the product was analyzed by a fluorescent DNA-sequencer (Applied Biosystems Inc.) to determine the base sequence of the cDNA 5'-terminal of about 400 bp. The sequence data were filed as a homo-protein cDNA bank data base.

### (3) Selection of cDNAs Encoding Proteins Having Secretory Signal Sequence

The base sequence registered in the homo-protein cDNA

bank was converted to three frames of amino acid sequences and the presence or absence of an open reading frame (ORF) beginning from the initiation codon. Then, the selection was made for the presence of a signal sequence that is characteristic to a secretory protein at the N-terminal of the portion encoded by ORF. These clones were sequenced from the both 5' and 3' directions by using the deletion method to determine the whole base sequence. The hydrophobicity/hydrophilicity profiles were obtained for proteins encoded by ORF by the Kyte-Doolittle method [Kyte, J. & Doolittle, R. F., J. Mol. Bio. 157: 105-132 (1982)] to examine the presence or absence of a hydrophobic region. In the case in which there is not a hydrophobic region of putative transmembrane domain(s) in the amino acid sequence of an encoded protein, this protein was considered as a membrane protein that did not possess a secretory protein or transmembrane domain(s).

#### (4) Construction of Secretory Signal Detection Vector pSSD3

One microgram of pSSD1 carrying the SV40 promoter and a cDNA encoding the protease domain of urokinase [Yokoyama-Kobayashi, M. et al., Gene 163: 193-196 (1995)] was digested with 5 units of BglII and 5 units of EcoRV. Then, after dephosphorylation at the 5' terminal by the CIP treatment, a DNA fragment of about 4.2 kbp was purified by cutting off from the gel of agarose gel electrophoresis.

Two oligo DNA linkers, L1 (5'-GATCCCGGGTCACGTGGGAT-3') and L2 (5'-ATCCACGTGACCCGG-3'), were synthesized and phosphorylated by T4 polynucleotide kinase. After annealing

of the both linkers, followed by ligation with the previously-prepared pSSD1 fragment by T4 DNA ligase, *Escherichia coli* JM109 was transformed. A plasmid pSSD3 was prepared from the transformant and the objective recombinant was confirmed by the determination of the base sequence of the linker-inserted fragment. Figure 1 illustrates the structure of the thus-obtained plasmid. The present plasmid vector carries three types of blunt-end formation restriction enzyme sites, SmaI, PmaCI, and EcoRV. Since these cleavage sites are positioned in succession at an interval of 7 bp, selection of an appropriate site in combination of three types of frames for the inserting cDNA allows to construct a vector expressing a fusion protein.

#### (5) Functional Verification of Secretory Signal Sequence

Whether the N-terminal hydrophobic region in the secretory protein clone candidate obtained in the above-mentioned steps functions as the secretory signal sequence was verified by the method described in the literature [Yokoyama-Kobayashi, M. et al., Gene 163: 193-196 (1995)]. First, the plasmid containing the target cDNA was cleaved at an appropriate restriction enzyme site that existed at the downstream from the portion expected for encoding the secretory signal sequence. In the case in which this restriction enzyme site was a protruding 5'-terminus, the site was blunt-ended by the Klenow treatment. Digestion with HindIII was further carried out and a DNA fragment containing the SV40 promoter and a cDNA encoding the secretory sequence at the downstream from the promoter was separated by agarose gel electrophoresis. This fragment was inserted between the

pSSD3 HindIII site and a restriction enzyme site selected so as to match with the urokinase-coding frame, thereby constructing a vector expressing a fusion protein of the secretory signal portion of the target cDNA and the urokinase protease domain (refer to Figure 2).

After *Escherichia coli* (host: JM109) bearing the fusion-protein expression vector was incubated at 37°C for 2 hours in 2 ml of the 2xYT culture medium containing 100 µg/ml ampicillin, the helper phage M13K07 (50 µl) was added and the incubation was continued at 37°C overnight. A supernatant separated by centrifugation underwent precipitation with polyethylene glycol to obtain single-stranded phage particles. These particles were suspended in 100 µl of 1 mM Tris-0.1 mM EDTA, pH 8 (TE). Also, there was used as a control a suspension of single-stranded particles prepared in the same manner from the vector pKA1-UPA containing pSSD3 and a full-length cDNA of urokinase [Yokoyama-Kobayashi, M. et al., Gene 163: 193-196 (1995)].

The simian-kidney-origin culture cells, COS7, were incubated at 37°C in the presence of 5% CO<sub>2</sub> in the Dulbecco's modified Eagle's culture medium (DMEM) containing 10% fetal calf albumin. Into a 6-well plate (Nunc Inc., 3 cm in the well diameter) were inoculated  $1 \times 10^5$  COS7 cells and incubation was carried out at 37°C for 22 hours in the presence of 5% CO<sub>2</sub>. After the culture medium was removed, the cell surface was washed with a phosphate buffer solution and then washed again with DMEM containing 50 mM Tris-hydrochloric acid (pH 7.5) (TDMEM). To the cells were added 1 µl of the single-stranded phage suspension, 0.6 ml of the

DMEM culture medium, and 3  $\mu$ l of TRANSFECTAM<sup>TM</sup> (IBF Inc.) and the resulting mixture was incubated at 37°C for 3 hours in the presence of 5% CO<sub>2</sub>. After the sample solution was removed, the cell surface was washed with TDMEM, 2 ml per well of DMEM containing 10% fetal calf albumin was added, and the incubation was carried out at 37°C for 2 days in the presence of 5% CO<sub>2</sub>.

To 10 ml of 50 mM phosphate buffer solution (pH 7.4) containing 2% bovine fibrinogen (Miles Inc.), 0.5% agarose, and 1 mM potassium chloride were added 10 units of human thrombin (Mochida Pharmaceutical Co., Ltd.) and the resulting mixture was solidified in a plate of 9 cm in diameter to prepare a fibrin plate. Ten microliters of the culture supernatant of the transfected COS7 cells were spotted on the fibrin plate, which was incubated at 37°C for 15 hours. The diameter of the thus-obtained clear circle was taken as an index for the urokinase activity. Table 2 shows the restriction enzyme site used for cutting off the cDNA fragment from each clone, the restriction enzyme site used for cleavage of pSSD3, and the presence or absence of a clear circle. Except for pSSD3 used as the control, each of the samples formed a clear circle to identify that urokinase was secreted in the culture medium. That is to say, it is indicated that each of the cDNA fragments codes for the amino acid sequence that functions as the secretory signal sequence.



Table 2

HP Number	Restriction Enzyme Site		Clear Circle
	cDNA*	Vector	
HP00658	HindIII (K)	SmaI	+
HP00714	PvuII	PmaCI	+
HP00876	NcoI (K)	PmaCI	+
HP01134	PmaCI	PmaCI	+
HP10029	ApaI (K)	SmaI	+
HP10189	BglI (K)	PmaCI	+
HP10269	PvuII	PmaCI	+
HP10298	HindIII (K)	PmaCI	+
HP10368	EcoRV	PmaCI	+
pKA1-UPA			+
pSSD3			-

\* (K) means that cleavage with the restriction enzyme is followed by the Klenow treatment.

#### (6) Protein Synthesis by In Vitro Translation

The plasmid vector carrying the cDNA of the present invention was utilized for the in vitro transcription/translation by the T<sub>N</sub>T rabbit reticulocyte lysate kit (Promega Biotec). In this case, [<sup>35</sup>S]methionine was added and the expression product was labeled with the radioisotope. All reactions were carried out by following the protocols attached to the kit. Two micrograms of the plasmid was allowed to react at 30°C for 90 minutes in total 25 ml of a reaction solution containing 12.5 µl of the T<sub>N</sub>T rabbit reticulocyte lysate, 0.5 µl of the buffer solution (attached to the kit), 2 µl of an amino acid mixture (methionine-free),

2  $\mu$ l (0.37 MBq/ $\mu$ l) of [ $^{35}$ S]methionine (Amersham Corporation), 0.5  $\mu$ l of T7 RNA polymerase, and 20 U of RNasin. Also, the experiment in the presence of the membrane system was carried out by adding 2.5  $\mu$ l of the dog pancreatic microsomal fraction (Promega Biotec) into this reaction system. To 3  $\mu$ l of the reaction solution was added 2  $\mu$ l of an SDS sampling buffer (125 mM Tris-hydrochloric acid buffer solution, pH 6.8, 120 mM 2-mercaptoethanol, 2% SDS solution, 0.025% bromophenol blue, and 20% glycerol) and the resulting solution was heated at 95°C for 3 minutes and then subjected to SDS-polyacrylamide gel electrophoresis. The molecular weight of the translation product was determined by carrying out the autoradiography. Table 3 shows the molecular weight of the in vitro translation product obtained from each of the clones in the presence/absence of the membrane microsomal fraction together with the calculated value of the molecular weight of the protein encoded by ORF of the cDNA.

Table 3

Se- quence No.	HP Number	Calcu- lated (Da)	In Vitro Translation Product (KDa)	
			Without Membrane System Added	With Membrane System Added*
1	HP00658	17,037	18	16
2	HP00714	37,106	47	-
3	HP00876	18,230	18	-
4	HP01134	42,947	42	49
5	HP10029	18,894	21	18
6	HP10189	9,113	12	-
7	HP10269	129,572	130	-
8	HP10298	13,161	16	-
9	HP10368	19,979	19	18

\* - means "Not examined".

#### (7) Clone Examples

<HP00658> (Sequence Number 1, 10, 19)

Determination of the whole base sequence for the cDNA insert of clone HP00658 obtained from the human fibrosarcoma cell line HT-1080 cDNA libraries revealed the structure consisting of a 5'-non-translation region of 55 bp, an ORF of 465 bp, and a 3'-non-translation region of 776 bp. The ORF codes for a protein consisting of 154 amino acid residues with a hydrophobic region of a putative secretory signal sequence at the N-terminal. Figure 3 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. Search of the protein data base using the amino acid sequence encoded by the ORF

revealed that the N-terminal 63 amino acid residues thereof were completely identical with those in the RANTES protein (EMBL Accession No. 21121) except for one amino acid residue at position 7 (arginine in RANTES and alanine in the present protein), but the sequences in both proteins were completely different after position 64. Hereupon, RANTES consisted of 91 amino acid residues, whereas the present protein consisted of longer 154 amino acid residues. The in vitro translation resulted in the formation of a translation product of 18 kDa that was almost consistent with the molecular weight of 17,037 predicted from the ORF. In this case, the addition of the microsome resulted in the formation of a 16-kDa product in which the secretory signal sequence portion was putatively removed by cleavage. This result together with the result on pSSD3 verifies that the present protein possesses the secretory signal sequence. Application of the (-3,-1) rule, a method for predicting the signal sequence cleavage site [von Heijne, G., Nucl. Acid Res. 14: 4683-4690 (1986)], allows to expect that the maturation protein starts from serine at position 24.

Comparison of the base sequences for the both proteins revealed that the base sequence from position 2 to position 325 in the present cDNA was deficient in the RANTES cDNA. It is considered that this deficiency resulted in induction of a frame shift to form an ORF of a different size. Some mutations were observed in other regions, wherein the homology was 97.7% up to position 241 and was 98.0% after position 325. RANTES has been obtained as a T cell-specific protein [Schall, T. J. et al., J. Immunol. 141: 1018-1025

(1988)], whereas the present cDNA was obtained from the fibrosarcoma cells. Accordingly, the present protein is considered to possess a different function from that of RANTES.

Furthermore, the search of GenBank using the base sequence of the present cDNA revealed that any EST possessing the homology of 90% or more was not found.

<HP00714> (Sequence Number 2, 11, 20)

Determination of the whole base sequence for the cDNA insert of clone HP00714 obtained from the human epidermoid carcinoma cell line KB cDNA libraries revealed the structure consisting of a 5'-non-translation region of 56 bp, an ORF of 948 bp, and a 3'-non-translation region of 2310 bp. The ORF codes for a protein consisting of 315 amino acid residues with a hydrophobic region of a putative secretory signal sequence at the N-terminal. Figure 4 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. The in vitro translation resulted in the formation of a translation product of 47 kDa that was somewhat larger than the molecular weight of 37,106 predicted from the ORF. Since the molecular weight of the human reticulocalbin analogous to the present protein is also larger by about 10 kDa than the molecular weight expected from the translation-product band on SDS-PAGE [Ozawa, M., J. Biochem. 117: 1113-1119 (1995)], the molecular weight difference in the present protein is considered to be arisen from its physicochemical properties. Application of the (-3,-1) rule, a method for predicting the signal sequence cleavage site, allows to expect that the maturation protein

starts from lysine at position 20. There is a possibility that the present protein exists in the endoplasmic reticulum because this protein possesses the C-terminal sequence HDEF analogous to KDEL, the signal motif sequence localized in the endoplasmic reticulum.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was analogous to the human reticulocalbin (GenBank Accession No. D42073). Table 4 indicates the comparison of the amino acid sequences between the human protein of the present invention (HP) and the human reticulocalbin (RC). - represents a gap, \* represents an amino acid residue identical to that in the protein of the present invention, and . represents an amino acid residue analogous to that in the protein of the present invention. The both proteins possessed a homology of 60.5%.

Table 4

[illegible]

\*\*\*.\*\*\*.\*\*\*\*\* \*\*\*\*\* \*.\*\*\*.\*\*\*.\*..\*\*\*\*\* \*\*\*\*\*,\*\*\*\*\*.\*\*\*\*\*.\*\*\*.\*

\*\*\*\*\*. . . . . \*\*\*\*\*. . \*\*. \*. . \*\*\*.

RC TKEEILENWNMFVGSQATNYGEDLTKNHDEL

Reticulocalbin is a protein localized on the membrane surface of the endoplasmic reticulum and has been considered to participate in the protein folding. Accordingly, the protein of the present invention is considered to be applicable to the folding process of recombinant proteins.

Determination of the whole base sequence for the cDNA insert of clone HP0876 obtained from the human stomach cancer cDNA libraries revealed the structure consisting of a 5'-non-translation region of 146 bp, an ORF of 477 bp, and a 3'-non-translation region of 529 bp. The ORF codes for a protein consisting of 158 amino acid residues with a hydrophobic region of a putative secretory signal sequence at the N-terminal. Figure 5 depicts the hydrophobicity/hydrophilicity

profile of the present protein obtained by the Kyte-Doolittle method. The in vitro translation resulted in the formation of a translation product of 18 kDa that was almost consistent with the molecular weight of 18,230 predicted from the ORF. In this case, the addition of the microsome resulted in the formation of a 16-kDa product in which the secretory signal sequence portion was putatively removed by cleavage. This result together with the result on pSSD3 verifies that the present protein possesses the secretory signal. Application of the (-3,-1) rule, a method for predicting the signal sequence cleavage site, allows to expect that the maturation protein starts from glycine at position 18 or aspartic acid at position 23.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was analogous to several type-C lectins. As an example, Table 5 indicates the comparison of the amino acid sequences between the human protein of the present invention (HP) and the rattlesnake lectin (CL) (Swiss-PROT Accession No. P21963). - represents a gap, \* represents an amino acid residue identical to that in the protein of the present invention, and . represents an amino acid residue analogous to that in the protein of the present invention. The both proteins possessed a homology of 35.3%.



Table 5

---

HP MASRSMRLLLLLSCLAKTGVLDIIMRPSAPGWFYHKSNCYGYFRKLRNWSDAEBCQS  
 .\*. \*. .. \*\* \*..\*..\*..\*..\*..\*.  
 CL NNCPLDWLPMNGLCYKIFNQLKTWEDAEMFCRK  
 HP YGNGAHLASILSLKEASTIAEYISGYQRSQ-PIWIGLHDPQKRQQWQWIDGAMYLYRSWS  
 \* \* \*\*\*\*. . \*. .\*\*\*\*\*.\*..\* .\*\*\*\*\*.\* \*..\* \*..\*.  
 CL YKPGCHLASPHRYGESLEIAEYISDYHKGQENVWIGLRDKKKDFSWEWTDORSCTDYLTWD  
 HP GKSMGG--NKH-CAEMSSNNFLTWSSNECNKRQHFLCKYRP  
 . . \*\*. \*. \* \*... \*... \*.... \*\*\*..  
 CL KNQPDHYQNKBFCEVLVSLTGYRLWNDQVCESKDAFLCQCKF

---

Furthermore, the search of GenBank using the base sequence of the present cDNA revealed that any EST possessing the homology of 90% or more was not found.

After 1 µg of the plasmid pHP00876 was digested with 20 units of PvuII, the product was subjected to 1% agarose gel electrophoresis and an about 700-bp DNA fragment was cut off from the gel. Next, 1 µg of pET-21a (Novagen) was digested with 20 units of NheI, the product was subjected to the Klenow treatment followed by 1% agarose gel electrophoresis and an about 5.4-kbp DNA fragment was cut off from the gel. After ligation of the vector fragment and the cDNA fragment using a ligation kit, *Escherichia coli* BL21 (DE3) (Novagen) was transformed. A plasmid pET876 was prepared from the transformant and the objective recombinant was confirmed from the restriction enzyme cleavage map. The present expression vector expresses a protein in which methionine-alanine was

inserted before a protein starting from serine at position 29 in the protein encoded by the clone HP00876.

A suspension of pET876/BL21 (DE3) in 5 ml of the LB culture medium containing 100 µg/ml ampicillin was incubated in a shaker at 37°C and isopropylthiogalactoside was added to make 1 mM when  $A_{600}$  reached to about 0.5. After the incubation was continued at 37°C for 6 hours, cells were collected by centrifugation and suspended in 25 ml of a column buffer solution for the amylose column (10 mM Tris-hydrochloric acid, pH 7.4, 200 mM NaCl, and 1 mM EDTA). The resulting suspension was sonicated and then the insoluble fraction was subjected to SDS-polyacrylamide electrophoresis to identify a band originating from the expression of the present vector at a position of about 14 kDa.

Since lectins recognize and then bind to sugar chains, lectins are useful as sugar-chain detection reagents and as affinity carriers for purification of glycoproteins. In addition, extracellular secretory lectins play important roles also in intercellular signal transduction and thereby are useful as medicines.

<HP01134> (Sequence Number 4, 13, 22)

Determination of the whole base sequence for the cDNA insert of clone HP01134 obtained from the human liver cDNA libraries revealed the structure consisting of a 5'-non-translation region of 116 bp, an ORF of 1131 bp, and a 3'-non-translation region of 502 bp. The ORF codes for a protein consisting of 376 amino acid residues with a hydrophobic region of a putative secretory signal sequence at the N-terminal. Figure 6 depicts the hydrophobicity/hydrophilicity

profile of the present protein obtained by the Kyte-Doolittle method. The in vitro translation resulted in the formation of a translation product of 42 kDa that was almost consistent with the molecular weight of 42,947 predicted from the ORF. In this case, the addition of the microsomes resulted in the formation of a 49-kDa product in which a sugar chain was putatively added by N-glycosylation after the secretion. Hereupon, there exist in the amino acid sequence of this protein four possible N-glycosylation sites (Asn-Gly-Thr at position 91, Asn-Glu-Thr at position 167, Asn-Thr-Ser at position 263, and Asn-Lys-Thr at position 272). The above result together with the result on pSSD3 verifies that the present protein possesses the secretory signal. Application of the (-3,-1) rule, a method for predicting the signal sequence cleavage site, allows to expect that the maturation protein starts from alanine at position 17 or valine at position 18.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was analogous to several cysteine proteinases. As an example, Table 6 indicates the comparison of the amino acid sequences between the human protein of the present invention (HP) and the tangerine cysteine proteinase (CP) (GenBank Accession No. Z47793). - represents a gap, \* represents an amino acid residue identical to that in the protein of the present invention, and . represents an amino acid residue analogous to that in the protein of the present invention. The both proteins possessed a homology of 49% among the N-terminal region of 286 amino acid residues.

Table 6

---

HP MVWKVAVFLSVALGIGAVPIDDPEDGGKH

\* \* \* \* \*

CP MTRLASGVLITLLVALAGIADGSRDIAGDILKLPSEAYRFFHNGGGGAKVNDDDDSVGTR

HP WVIVAGSNGWYNYRHQADACHAYQIIHRNGIPDEQIVVMYDDIAYSEDNPTPGIVINR

\* . \* . \* \* \* \* . \* \* \* \* \* \* \* \* . . . . \* . \* \* . \* . \* \* \* \* \* . \* . \* \* . \* .

CP WAVLLAGSNGFWNYRHQADICHAYQLLRKGGKLDENIIVFMYDDIAFNEENPRPGVIINH

HP PNGTDVYQGVPKDYGEDVTPQNFLAVLRGDAEAVKGIGSGKVLKSGPQDHVFIYFTDHG

\* . \* . \* \* \* . \* \* \* \* \* \* \* \* \* . . . \* . \* . \* . \* \* \* \* \* . \* \* . \* . \* . \* \* .

CP PHGDDVYKGVPKDYGEDVTVBKFFAVVLGNKTALTG-GSGKVVDSGPNDHIFIRYSDHG

HP STGILVFPNED-LHVKDLNETIHYMYKHMYKRMVFIYBACBSGSMN-HLPDINIVYAT

. . \* . \* . \* . . . . \* . . . . . \* . . . \* \* \* \* \* \* \* . . \* . . . \* . \* \* \*

CP GPGVLGMPTSRYYADELIDVLKKKHASGNYKSLVFYLBACBSGSIFEGLLLBGLNIYAT

HP TAANPRESSYACY--DEKRSTY--LGDWYSVNWMESSDVEDLTKETLHKQYHLVKS

\* \* . \* . \* \* . \* . . . . \* \* \* \* \* . \* \* \* \* \* . \* . \* \* \* . \* . \* \* .

CP TASNABBSSWGTYCPGEIPGPPPEYSTCLGDLYSIAWMEDSDIHNLRTEHLHQYELVKT

HP HT-----NTSHVMQYGNKTIISTMKVMQFQGMKRKASSPVPLPPVTHLDLTPSPDVPLTIM

. \* . \* \* \* \* \* . . \* . . . . \*

CP RTASYSYSGSHVMQYGDIGLSKNNLFTYLGTNPANDNYTFVDENSLRPASKAVNQROADL

---

Furthermore, the search of GenBank using the base sequence of the present cDNA revealed that there existed some ESTs possessing the homology of 90% or more (for example, Accession No. F01300), but they were shorter than the present cDNA and any molecule containing the initiation codon was not identified.

Extracellular secretory proteases possess a variety of physiological functions and thereby are useful as medicines. In addition, the proteases have been utilized as research reagents for the structure analysis of proteins by restricted degradation and so on.

<HP10029> (Sequence Number 5, 14, 23)

Determination of the whole base sequence for the cDNA insert of clone HP10029 obtained from the human epidermoid carcinoma cell line KB cDNA libraries revealed the structure consisting of a 5'-non-translation region of 8 bp, an ORF of 522 bp, and a 3'-non-translation region of 458 bp. The ORF codes for a protein consisting of 173 amino acid residues with a hydrophobic region of a putative secretory signal sequence at the N-terminal. Figure 7 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. The in vitro translation resulted in the formation of a translation product of 21 kDa that was almost consistent with the molecular weight of 18,894 predicted from the ORF. In this case, the addition of the microsome resulted in the formation of a 18-kDa product in which the secretory signal sequence portion was putatively removed by cleavage. This result together with the result on pSSD3 verifies that the present protein possesses the secretory signal sequence. Application of the (-3,-1) rule, a method for predicting the signal sequence cleavage site, allows to expect that the maturation protein starts from valine at position 32. There is a possibility that the present protein exists in the endoplasmic reticulum because this protein possesses the C-

terminal sequence RTEL analogous to KDEL, the signal motif sequence localized in the endoplasmic reticulum.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was not homologous with any of known proteins. Hereupon, the search of GenBank using the base sequence revealed that there existed some ESTs possessing the homology of 90% or more (for example, Accession No. H87021), but they were shorter than the present cDNA and any molecule containing the initiation codon was not identified.

<HP10189> (Sequence Number 6, 15, 24)

Determination of the whole base sequence for the cDNA insert of clone HP10189 obtained from the human epidermoid carcinoma cell line KB cDNA libraries revealed the structure consisting of a 5'-non-translation region of 101 bp, an ORF of 222 bp, and a 3'-non-translation region of 67 bp. The ORF codes for a protein consisting of 73 amino acid residues with a hydrophobic region of a putative secretory signal sequence at the N-terminal. Figure 8 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. The in vitro translation resulted in the formation of a translation product of 10 kDa that was almost consistent with the molecular weight of 9,113 predicted from the ORF. Application of the (-3,-1) rule, a method for predicting the signal sequence cleavage site, allows to expect that the maturation protein starts from alanine at position 27.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was

not homologous with any of known proteins. Hereupon, the search of GenBank using the base sequence revealed that there existed some ESTs possessing the homology of 90% or more and containing the initiation codon (for example, Accession No. N56270), but a frame shift had occurred and the same ORF as that in the present cDNA was not identified.

<HP10269> (Sequence Number 7, 16, 25)

Determination of the whole base sequence for the cDNA insert of clone HP10269 obtained from the human lymphoma cell line U937 cDNA libraries revealed the structure consisting of a 5'-non-translation region of 753 bp, an ORF of 351 bp, and a 3'-non-translation region of 395 bp. The ORF codes for a protein consisting of 1172 amino acid residues with a hydrophobic region of a putative secretory signal sequence at the N-terminal. Figure 9 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. The in vitro translation resulted in the formation of a translation product of 130 kDa that was almost consistent with the molecular weight of 129,571 predicted from the ORF. Application of the (-3,-1) rule, a method for predicting the signal sequence cleavage site, allows to expect that the maturation protein starts from glutamine at position 18.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was analogous to the B3 chain of laminin S. Table 7 indicates the comparison of the amino acid sequences between the human protein of the present invention (HP) and the B3 chain of human laminin S (B3) (GenBank Accession No. L25541)

Table 7

Amino Acid Residue Number	HP	B3
124	Gln	Arg
269	Pro	Deficient
388	Pro	Ala
426	Gln	Arg
427	Gly	Arg
439	Arg	Deficient
441	Asp	Glu
603	Arg	Pro
815	Gly	Ala

Comparison of the base sequence of the present cDNA and the base sequence described in the data base reveals that the 5'-terminus in the present cDNA is longer by 600 or more bp and the 81-bp 5'-terminus in the base sequence described in the data base is not consistent at all with the base sequence of the present cDNA. Accordingly, the both proteins originate from different mRNAs.

As an extracellular matrix, laminin deeply participates in the proliferation and differentiation of cells. Accordingly, laminin has been employed as an additive for the cell culture and so on.

<HP10298> (Sequence Number 8, 17, 26)

Determination of the whole base sequence for the cDNA insert of clone HP10298 obtained from the human stomach cancer cDNA libraries revealed the structure consisting of a 5'-non-translation region of 137 bp, an ORF of 369 bp, and a



3'-non-translation region of 580 bp. The ORF codes for a protein consisting of 122 amino acid residues with a hydrophobic region of a putative secretory signal sequence at the N-terminal. Figure 10 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. The in vitro translation resulted in the formation of a translation product of 16 kDa that was almost consistent with the molecular weight of 13,161 predicted from the ORF. Application of the (-3,-1) rule, a method for predicting the signal sequence cleavage site, allows to expect that the maturation protein starts from leucine at position 18. There is also a possibility that the present protein possessing the hydrophobic C-terminal sequence of about 20 amino acid residues binds to the membrane via this portion.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was not homologous with any of known proteins. Hereupon, the search of GenBank using the base sequence revealed that there existed some ESTs possessing the homology of 90% or more and containing the initiation codon (for example, Accession No. D78655), but many sequences were not distinct and the same ORF as that in the present cDNA was not identified.

<HP10368> (Sequence Number 9, 18, 27)

Determination of the whole base sequence for the cDNA insert of clone HP10368 obtained from the human stomach cancer cDNA libraries revealed the structure consisting of a 5'-non-translation region of 72 bp, an ORF of 528 bp, and a 3'-non-translation region of 266 bp. The ORF codes for a

protein consisting of 175 amino acid residues with a hydrophobic region of a putative secretory signal sequence at the N-terminal. Figure 11 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. The in vitro translation resulted in the formation of a translation product of 20 kDa that was almost consistent with the molecular weight of 19,979 predicted from the ORF. In this case, the addition of the microsome resulted in the formation of a 19-kDa product in which the secretory signal sequence portion was putatively removed by cleavage. This result together with the result on pSSD3 verifies that the present protein possesses the secretory signal. Application of the (-3,-1) rule, a method for predicting the signal sequence cleavage site, allows to expect that the maturation protein starts from leucine at position 19 or arginine at position 21. There is a possibility that the present protein exists in the endoplasmic reticulum because this protein possesses the C-terminal sequence KTEL analogous to KDEL, the signal motif sequence localized in the endoplasmic reticulum.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was not homologous with any of known proteins. Hereupon, the search of GenBank using the base sequence revealed that there existed some ESTs possessing the homology of 90% or more and containing the initiation codon (for example, Accession No. T86663), but many sequences were not distinct and the same ORF as that in the present cDNA was not identified.

#### INDUSTRIAL APPLICATION

The present invention provides human proteins having secretory signal sequences and cDNAs encoding said proteins. All of the proteins of the present invention are putative proteins controlling the proliferation and differentiation of the cells, because said proteins are secreted outside the cells and exist in the extracellular liquid or on the cell membrane surface. Therefore, the proteins of the present invention can be used as pharmaceuticals or as antigens for preparing antibodies against said proteins. Furthermore, said DNAs can be used for the expression of large amounts of said proteins.

In addition to the activities and uses described above, the polynucleotides and proteins of the present invention may exhibit one or more of the uses or biological activities (including those associated with assays cited herein) identified below. Uses or activities described for proteins of the present invention may be provided by administration or use of such proteins or by administration or use of polynucleotides encoding such proteins (such as, for example, in gene therapies or vectors suitable for introduction of DNA).

#### Research Uses and Utilities

The polynucleotides provided by the present invention can be used by the research community for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a

particular stage of tissue differentiation or development or in disease states); as molecular weight markers on Southern gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; as a probe to "subtract-out" known sequences in the process of discovering other novel polynucleotides; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination of expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris et al., Cell 75:791-803 (1993)) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

The proteins provided by the present invention can similarly be used in assay to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers

for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Where the protein binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the protein can be used to identify the other protein with which binding occurs or to identify inhibitors of the binding interaction. Proteins involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E.F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S.L. and A.R. Kimmel eds., 1987.

#### Nutritional Uses

Polynucleotides and proteins of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases

the protein or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the protein or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

Cytokine and Cell Proliferation/Differentiation Activity

A protein of the present invention may exhibit cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of a protein of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+ (preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e and CMK.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for T-cell or thymocyte proliferation include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A.M. Kruisbeek, D.H. Margulies, E.M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays

for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Bertagnolli et al., J. Immunol. 145:1706-1712, 1990; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Bertagnolli, et al., J. Immunol. 149:3778-3783, 1992; Bowman et al., J. Immunol. 152: 1756-1761, 1994.

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described in: Polyclonal T cell stimulation, Kruisbeek, A.M. and Shevach, E.M. In Current Protocols in Immunology. J.E.e.a. Coligan eds. Vol 1 pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto. 1994; and Measurement of mouse and human Interferon  $\gamma$ , Schreiber, R.D. In Current Protocols in Immunology. J.E.e.a. Coligan eds. Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described in: Measurement of Human and Murine Interleukin 2 and Interleukin 4, Bottomly, K., Davis, L.S. and Lipsky, P.E. In Current Protocols in Immunology. J.E.e.a. Coligan eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries et al., J. Exp. Med. 173:1205-1211, 1991; Moreau et al., Nature 336:690-692, 1988; Greenberger et al., Proc. Natl. Acad. Sci. U.S.A. 80:2931-2938, 1983; Measurement of mouse and human interleukin 6 -Nordan, R. In Current Protocols in Immunology. J.E.e.a. Coligan eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; Smith et al., Proc. Natl. Acad. Sci.

U.S.A. 83:1857-1861, 1986; Measurement of human Interleukin 11 - Bennett, F., Giannotti, J., Clark, S.C. and Turner, K. J. In Current Protocols in Immunology. J.E.e.a. Coligan eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto. 1991; Measurement of mouse and human Interleukin 9 - Ciarletta, A., Giannotti, J., Clark, S.C. and Turner, K.J. In Current Protocols in Immunology. J.E.e.a. Coligan eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A.M. Kruisbeek, D.H. Margulies, E.M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function; Chapter 6, Cytokines and their cellular receptors; Chapter 7, Immunologic studies in Humans); Weinberger et al., Proc. Natl. Acad. Sci. USA 77:6091-6095, 1980; Weinberger et al., Eur. J. Immun. 11:405-411, 1981; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988.

#### Immune Stimulating or Suppressing Activity

A protein of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency (SCID)), e.g., in regulating



(up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases caused by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpesviruses, mycobacteria, Leishmania spp., malaria spp. and various fungal infections such as candidiasis. Of course, in this regard, a protein of the present invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein of the present invention may also be useful in the treatment of allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein of the present invention.

Using the proteins of the invention it may also be

possible to immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as , for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a molecule which inhibits or blocks interaction of a B7 lymphocyte antigen with its natural

ligand(s) on immune cells (such as a soluble, monomeric form of a peptide having B7-2 activity alone or in conjunction with a monomeric form of a peptide having an activity of another B lymphocyte antigen (e.g., B7-1, B7-3) or blocking antibody), prior to transplantation can lead to the binding of the molecule to the natural ligand(s) on the immune cells without transmitting the corresponding costimulatory signal. Blocking B lymphocyte antigen function in this matter prevents cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, the lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular blocking reagents in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins in vivo as described in Lenschow et al., *Science* 257:789-792 (1992) and Turka et al., *Proc. Natl. Acad. Sci USA*, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., *Fundamental Immunology*, Raven Press, New York, 1989, pp. 846-847) can be used to determine

the effect of blocking B lymphocyte antigen function in vivo on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block costimulation of T cells by disrupting receptor:ligand interactions of B lymphocyte antigens can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythmatosis in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (preferably a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy.

Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response through stimulating B lymphocyte antigen function may be useful in cases of viral infection. In addition, systemic viral diseases such as influenza, the common cold, and encephalitis might be alleviated by the administration of stimulatory forms of B lymphocyte antigens systemically.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells in vitro with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the in vitro activated T cells into the patient. Another method of enhancing anti-viral immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells in vivo.

In another application, up regulation or enhancement of antigen function (preferably B lymphocyte antigen function) may be useful in the induction of tumor immunity. Tumor cells (e.g., sarcoma, melanoma, lymphoma, leukemia, neuroblastoma, carcinoma) transfected with a nucleic acid

encoding at least one peptide of the present invention can be administered to a subject to overcome tumor-specific tolerance in the subject. If desired, the tumor cell can be transfected to express a combination of peptides. For example, tumor cells obtained from a patient can be transfected ex vivo with an expression vector directing the expression of a peptide having B7-2-like activity alone, or in conjunction with a peptide having B7-1-like activity and/or B7-3-like activity. The transfected tumor cells are returned to the patient to result in expression of the peptides on the surface of the transfected cell. Alternatively, gene therapy techniques can be used to target a tumor cell for transfection in vivo.

The presence of the peptide of the present invention having the activity of a B lymphocyte antigen(s) on the surface of the tumor cell provides the necessary costimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient amounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I  $\alpha$  chain protein and  $\beta_2$  microglobulin protein or an MHC class II  $\alpha$  chain protein and an MHC class II  $\beta$  chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T

cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A.M. Kruisbeek, D.H. Margulies, E.M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564-1572, 1985; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564-1572, 1985; Takai et al., J. Immunol. 137:3494-3500, 1986; Bowman et al., J. Virology 61:1992-1998; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., Cellular Immunology 133:327-341, 1991;

Brown et al., J. Immunol. 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, J. Immunol. 144:3028-3033, 1990; and Assays for B cell function: In vitro antibody production, Mond, J.J. and Brunswick, M. In Current Protocols in Immunology. J.E.e.a. Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto. 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A.M. Kruisbeek, D.H. Margulies, E.M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., J. Immunol. 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery et al., J. Immunol. 134:536-544, 1995; Inaba et al., Journal of Experimental Medicine 173:549-559, 1991; Macatonia et al., Journal of Immunology 154:5071-5079, 1995; Porgador et al., Journal of Experimental Medicine 182:255-260, 1995; Nair et al., Journal of Virology



67:4062-4069, 1993; Huang et al., Science 264:961-965, 1994; Macatonia et al., Journal of Experimental Medicine 169:1255-1264, 1989; Bhardwaj et al., Journal of Clinical Investigation 94:797-807, 1994; and Inaba et al., Journal of Experimental Medicine 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz et al., Cytometry 13:795-808, 1992; Gorczyca et al., Leukemia 7:659-670, 1993; Gorczyca et al., Cancer Research 53:1945-1951, 1993; Itoh et al., Cell 66:233-243, 1991; Zacharchuk, Journal of Immunology 145:4037-4045, 1990; Zamai et al., Cytometry 14:891-897, 1993; Gorczyca et al., International Journal of Oncology 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica et al., Blood 84:111-117, 1994; Fine et al., Cellular Immunology 155:111-122, 1994; Galy et al., Blood 85:2770-2778, 1995; Toki et al., Proc. Nat. Acad Sci. USA 88:7548-7551, 1991.

#### Hematopoiesis Regulating Activity

A protein of the present invention may be useful in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell deficiencies. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in

combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either in-vivo or ex-vivo (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson et al. Cellular Biology 15:141-151, 1995; Keller et al., Molecular and Cellular Biology 13:473-486, 1993; McClanahan et al., Blood 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, M.G. In Culture of Hematopoietic Cells. R.I. Freshney, et al. eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, NY. 1994; Hirayama et al., Proc. Natl. Acad. Sci. USA 89:5907-5911, 1992; Primitive hematopoietic colony forming cells with high proliferative potential, McNiece, I.K. and Briddell, R.A. In Culture of Hematopoietic Cells. R.I. Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, NY. 1994; Neben et al., Experimental Hematology 22:353-359, 1994; Cobblestone area forming cell assay, Ploemacher, R.E. In Culture of Hematopoietic Cells. R.I. Freshney, et al. eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, NY. 1994; Long term bone marrow cultures in the presence of stromal cells, Spooncer, E., Dexter, M. and Allen, T. In Culture of Hematopoietic Cells. R.I. Freshney, et al. eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, NY. 1994; Long term culture initiating cell assay, Sutherland, H.J. In Culture of Hematopoietic Cells. R.I. Freshney, et al. eds. Vol pp. 139-162, Wiley-Liss, Inc., New York, NY. 1994.

Tissue Growth Activity

A protein of the present invention also may have utility in compositions used for bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as for wound healing and tissue repair and replacement, and in the treatment of burns, incisions and ulcers.

A protein of the present invention, which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Such a preparation employing a protein of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A protein of this invention may also be used in the treatment of periodontal disease, and in other tooth repair processes. Such agents may provide an environment to attract bone-forming cells, stimulate growth of bone-forming cells or induce differentiation of progenitors of bone-forming cells. A protein of the invention may also be useful in the treatment of osteoporosis or osteoarthritis, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes.

Another category of tissue regeneration activity that may

be attributable to the protein of the present invention is tendon/ligament formation. A protein of the present invention, which induces tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide an environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors ex vivo for return in vivo to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The protein of the present invention may also be useful

for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a protein may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a protein of the invention.

Proteins of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

It is expected that a protein of the present invention may also exhibit activity for generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of

cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring to allow normal tissue to regenerate. A protein of the invention may also exhibit angiogenic activity.

A protein of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A protein of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium ).

Assays for wound healing activity include, without limitation, those described in: Winter, Epidermal Wound Healing, pps. 71-112 (Maibach, HI and Rovee, DT, eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978).

#### Activin/Inhibin Activity

A protein of the present invention may also exhibit activin- or inhibin-related activities. Inhibins are characterized by their ability to inhibit the release of

follicle stimulating hormone (FSH), while activins and are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a protein of the present invention, alone or in heterodimers with a member of the inhibin  $\alpha$  family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the protein of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin- $\beta$  group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, United States Patent 4,798,885. A protein of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as cows, sheep and pigs.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for activin/inhibin activity include, without limitation, those described in: Vale et al., Endocrinology 91:562-572, 1972; Ling et al., Nature 321:779-782, 1986; Vale et al., Nature 321:776-779, 1986; Mason et al., Nature 318:659-663, 1985; Forage et al., Proc. Natl. Acad. Sci. USA 83:3091-3095, 1986.

#### Chemotactic/Chemokinetic Activity

A protein of the present invention may have chemotactic



or chemokinetic activity (e.g., act as a chemokine) for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. Chemotactic and chemokinetic proteins can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic proteins provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to

another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: Current Protocols in Immunology, Ed by J.E. Coligan, A.M. Kruisbeek, D.H. Margulies, E.M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub et al. J. Clin. Invest. 95:1370-1376, 1995; Lind et al. APMIS 103:140-146, 1995; Muller et al Eur. J. Immunol. 25: 1744-1748; Gruber et al. J. of Immunol. 152:5860-5867, 1994; Johnston et al. J. of Immunol. 153: 1762-1768, 1994.

#### Hemostatic and Thrombolytic Activity

A protein of the invention may also exhibit hemostatic or thrombolytic activity. As a result, such a protein is expected to be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A protein of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke).

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assay for hemostatic and thrombolytic activity include, without limitation, those described in: Linet et al., J. Clin. Pharmacol. 26:131-140, 1986; Burdick et al., Thrombosis Res. 45:413-419, 1987; Humphrey et al., Fibrinolysis 5:71-79

(1991); Schaub, Prostaglandins 35:467-474, 1988.

#### Receptor/Ligand Activity

A protein of the present invention may also demonstrate activity as receptors, receptor ligands or inhibitors or agonists of receptor/ligand interactions. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses). Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in: Current Protocols in Immunology, Ed by J.E. Coligan, A.M. Kruisbeek, D.H. Margulies, E.M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions 7.28.1-7.28.22), Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein

et al., J. Exp. Med. 169:149-160 1989; Stoltenborg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995.

#### Anti-Inflammatory Activity

Proteins of the present invention may also exhibit anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response. Proteins exhibiting such activities can be used to treat inflammatory conditions including chronic or acute conditions), including without limitation inflammation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Proteins of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material.

#### Tumor Inhibition Activity

In addition to the activities described above for immunological treatment or prevention of tumors, a protein of the invention may exhibit other anti-tumor activities. A

protein may inhibit tumor growth directly or indirectly (such as, for example, via ADCC). A protein may exhibit its tumor inhibitory activity by acting on tumor tissue or tumor precursor tissue, by inhibiting formation of tissues necessary to support tumor growth (such as, for example, by inhibiting angiogenesis), by causing production of other factors, agents or cell types which inhibit tumor growth, or by suppressing, eliminating or inhibiting factors, agents or cell types which promote tumor growth

#### Other Activities

A protein of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent

behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

## SEQUENCE LISTING

Sequence No.: 1

Sequence length: 154

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Fibrosarcoma

Cell line: HT-1080

Clone name: HP00658

## Sequence description

Met Lys Val Ser Ala Ala Ala Leu Ala Val Ile Leu Ile Ala Thr Ala  
1 5 10 15  
Leu Cys Ala Pro Ala Ser Ala Ser Pro Tyr Ser Ser Asp Thr Thr Pro  
20 25 30  
Cys Cys Phe Ala Tyr Ile Ala Arg Pro Leu Pro Arg Ala His Ile Lys  
35 40 45  
Glu Tyr Phe Tyr Thr Ser Gly Lys Cys Ser Asn Pro Ala Val Val His  
50 55 60  
Arg Ser Arg Met Pro Lys Arg Glu Gly Gln Gln Val Trp Gln Asp Phe  
65 70 75 80  
Leu Tyr Asp Ser Arg Leu Asn Lys Gly Lys Leu Cys His Pro Lys Glu  
85 90 95  
Pro Pro Ser Val Cys Gln Pro Arg Glu Glu Met Gly Ser Gly Val His  
100 105 110  
Gln Leu Phe Gly Asp Glu Leu Gly Trp Arg Val Leu Glu Pro Glu Leu

65

115 120 125  
Thr Gln Ile Cys Leu Phe Leu Leu Ala Leu Val Leu Ala Trp Glu Ala  
130 135 140  
Ser Pro His Tyr Pro Thr Pro Pro Ala Pro  
145 150

Sequence No.: 2

Sequence length: 315

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Epidermoid carcinoma

Cell line: KB

Clone name: HP00714

Sequence description

Met Asp Leu Arg Gln Phe Leu Met Cys Leu Ser Leu Cys Thr Ala Phe  
1 5 10 15  
Ala Leu Ser Lys Pro Thr Glu Lys Lys Asp Arg Val His His Glu Pro  
20 25 30  
Gln Leu Ser Asp Lys Val His Asn Asp Ala Gln Ser Phe Asp Tyr Asp  
35 40 45  
His Asp Ala Phe Leu Gly Ala Glu Glu Ala Lys Thr Phe Asp Gln Leu  
50 55 60  
Thr Pro Glu Glu Ser Lys Glu Arg Leu Gly Lys Ile Val Ser Lys Ile  
65 70 75 80  
Asp Gly Asp Lys Asp Gly Phe Val Thr Val Asp Glu Leu Lys Asp Trp



66

	85	90	95
Ile Lys Phe Ala Gln Lys Arg Trp Ile Tyr Glu Asp Val Glu Arg Gln			
100	105	110	
Trp Lys Gly His Asp Leu Asn Glu Asp Gly Leu Val Ser Trp Glu Glu			
115	120	125	
Tyr Lys Asn Ala Thr Tyr Gly Tyr Val Leu Asp Asp Pro Asp Pro Asp			
130	135	140	
Asp Gly Phe Asn Tyr Lys Gln Met Met Val Arg Asp Glu Arg Arg Phe			
145	150	155	160
Lys Met Ala Asp Lys Asp Gly Asp Leu Ile Ala Thr Lys Glu Glu Phe			
165	170	175	
Thr Ala Phe Leu His Pro Glu Glu Tyr Asp Tyr Met Lys Asp Ile Val			
180	185	190	
Val Gln Glu Thr Met Glu Asp Ile Asp Lys Asn Ala Asp Gly Phe Ile			
195	200	205	
Asp Leu Glu Glu Tyr Ile Gly Asp Met Tyr Ser His Asp Gly Asn Thr			
210	215	220	
Asp Glu Pro Glu Trp Val Lys Thr Glu Arg Glu Gln Phe Val Glu Phe			
225	230	235	240
Arg Asp Lys Asn Arg Asp Gly Lys Met Asp Lys Glu Glu Thr Lys Asp			
245	250	255	
Trp Ile Leu Pro Ser Asp Tyr Asp His Ala Glu Ala Glu Ala Arg His			
260	265	270	
Leu Val Tyr Glu Ser Asp Gln Asn Lys Asp Gly Lys Leu Thr Lys Glu			
275	280	285	
Glu Ile Val Asp Lys Tyr Asp Leu Phe Val Gly Ser Gln Ala Thr Asp			
290	295	300	
Phe Gly Glu Ala Leu Val Arg His Asp Glu Phe			
305	310	315	

Sequence No.: 3

Sequence length: 158

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Stomach cancer

Clone name: HP00876

Sequence description

Met Ala Ser Arg Ser Met Arg Leu Leu Leu Leu Leu Ser Cys Leu Ala

1 5 10 15

Lys Thr Gly Val Leu Gly Asp Ile Ile Met Arg Pro Ser Cys Ala Pro

20 25 30

Gly Trp Phe Tyr His Lys Ser Asn Cys Tyr Gly Tyr Phe Arg Lys Leu

35 40 45

Arg Asn Trp Ser Asp Ala Glu Leu Glu Cys Gln Ser Tyr Gly Asn Gly

50 55 60

Ala His Leu Ala Ser Ile Leu Ser Leu Lys Glu Ala Ser Thr Ile Ala

65 70 75 80

Glu Tyr Ile Ser Gly Tyr Gln Arg Ser Gln Pro Ile Trp Ile Gly Leu

85 90 95

His Asp Pro Gln Lys Arg Gln Gln Trp Gln Trp Ile Asp Gly Ala Met

100 105 110

Tyr Leu Tyr Arg Ser Trp Ser Gly Lys Ser Met Gly Gly Asn Lys His

115 120 125

Cys Ala Glu Met Ser Ser Asn Asn Asn Phe Leu Thr Trp Ser Ser Asn

68

130 135 140  
Glu Cys Asn Lys Arg Gln His Phe Leu Cys Lys Tyr Arg Pro  
145 150 155

Sequence No.: 4

Sequence length: 376

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Liver

Clone name: HP01134

Sequence description

Met Val Trp Lys Val Ala Val Phe Leu Ser Val Ala Leu Gly Ile Gly  
1 5 10 15  
Ala Val Pro Ile Asp Asp Pro Glu Asp Gly Gly Lys His Trp Val Val  
20 25 30  
Ile Val Ala Gly Ser Asn Gly Trp Tyr Asn Tyr Arg His Gln Ala Asp  
35 40 45  
Ala Cys His Ala Tyr Gln Ile Ile His Arg Asn Gly Ile Pro Asp Glu  
50 55 60  
Gln Ile Val Val Met Met Tyr Asp Asp Ile Ala Tyr Ser Glu Asp Asn  
65 70 75 80  
Pro Thr Pro Gly Ile Val Ile Asn Arg Pro Asn Gly Thr Asp Val Tyr  
85 90 95  
Gln Gly Val Pro Lys Asp Tyr Thr Gly Glu Asp Val Thr Pro Gln Asn  
100 105 110

69

Phe Leu Ala Val Leu Arg Gly Asp Ala Glu Ala Val Lys Gly Ile Gly  
115 120 125

Ser Gly Lys Val Leu Lys Ser Gly Pro Gln Asp His Val Phe Ile Tyr  
130 135 140

Phe Thr Asp His Gly Ser Thr Gly Ile Leu Val Phe Pro Asn Glu Asp  
145 150 155 160

Leu His Val Lys Asp Leu Asn Glu Thr Ile His Tyr Met Tyr Lys His  
165 170 175

Lys Met Tyr Arg Lys Met Val Phe Tyr Ile Glu Ala Cys Glu Ser Gly  
180 185 190

Ser Met Met Asn His Leu Pro Asp Asn Ile Asn Val Tyr Ala Thr Thr  
195 200 205

Ala Ala Asn Pro Arg Glu Ser Ser Tyr Ala Cys Tyr Tyr Asp Glu Lys  
210 215 220

Arg Ser Thr Tyr Leu Gly Asp Trp Tyr Ser Val Asn Trp Met Glu Asp  
225 230 235 240

Ser Asp Val Glu Asp Leu Thr Lys Glu Thr Leu His Lys Gln Tyr His  
245 250 255

Leu Val Lys Ser His Thr Asn Thr Ser His Val Met Gln Tyr Gly Asn  
260 265 270

Lys Thr Ile Ser Thr Met Lys Val Met Gln Phe Gln Gly Met Lys Arg  
275 280 285

Lys Ala Ser Ser Pro Val Pro Leu Pro Pro Val Thr His Leu Asp Leu  
290 295 300

Thr Pro Ser Pro Asp Val Pro Leu Thr Ile Met Lys Arg Lys Leu Met  
305 310 315 320

Asn Thr Asn Asp Leu Glu Glu Ser Arg Gln Leu Thr Glu Glu Ile Gln  
325 330 335

Arg His Leu Asp Tyr Glu Tyr Ala Leu Arg His Leu Tyr Val Leu Val

70

```

          340                      345                      350
Asn Leu Cys Glu Lys Pro Tyr Pro Leu His Arg Ile Lys Leu Ser Met

          355                      360                      365
Asp His Val Cys Leu Gly His Tyr

          370                      375

```

**Sequence No.: 5**

Sequence length: 173

Sequence type: Amino acid

**Topology: Linear**

**Sequence kind: Protein**

**Hypothetical: No**

**Original source:**

Organism species: *Homo sapiens*

Cell kind: Epidermoid carcinoma

**Cell line:** KB

Clone name: HP10029

### Sequence description

Met Ala Ala Pro Ser Gly Gly Trp Asn Gly Val Arg Ala Ser Leu Trp

**1                      5                      10                      15**

Ala Ala Leu Leu Leu Gly Ala Val Ala Leu Arg Pro Ala Glu Ala Val

**20                      25                      30**

Ser Glu Pro Thr Thr Val Ala Phe Asp Val Arg Pro Gly Gly Val Val

**35                      40                      45**

His Ser Phe Ser His Asn Val Gly Pro Gly Asp Lys Tyr Thr Cys Met

**50                      55                      60**

Phe Thr Tyr Ala Ser Gln Gly Gly Thr Asn Glu Gln Trp Gln Met Ser

**65                      70                      75                      80**

Leu Gly Thr Ser Glu Asp His Gln His Phe Thr Cys Thr Ile Trp Arg

71  
85 90 95  
Pro Gln Gly Lys Ser Tyr Leu Tyr Phe Thr Gln Phe Lys Ala Glu Val  
100 105 110  
Arg Gly Ala Glu Ile Glu Tyr Ala Met Ala Tyr Ser Lys Ala Ala Phe  
115 120 125  
Glu Arg Glu Ser Asp Val Pro Leu Lys Thr Glu Glu Phe Glu Val Thr  
130 135 140  
Lys Thr Ala Val Ala His Arg Pro Gly Ala Phe Lys Ala Glu Leu Ser  
145 150 155 160  
Lys Leu Val Ile Val Ala Lys Ala Ser Arg Thr Glu Leu  
165 170

Sequence No.: 6

Sequence length: 73

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Epidermoid carcinoma

Cell line: KB

Clone name: HP10189

Sequence description

Met Gly Val Lys Leu Glu Ile Phe Arg Met Ile Ile Tyr Leu Thr Phe  
1 5 10 15  
Pro Val Ala Met Phe Trp Val Ser Asn Gln Ala Glu Trp Phe Glu Asp  
20 25 30  
Asp Val Ile Gln Arg Lys Arg Glu Leu Trp Pro Pro Glu Lys Leu Gln

72

35                                      40                                      45  
 Glu Ile Glu Glu Phe Lys Glu Arg Leu Arg Lys Arg Arg Glu Glu Lys  
 50                                      55                                      60  
 Leu Leu Arg Asp Ala Gln Gln Asn Ser  
 65                                      70

Sequence No.: 7

Sequence length: 1172

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Histiocyte lymphoma

Cell line: U937

Clone name: HP10269

Sequence description

Met Arg Pro Phe Phe Leu Leu Cys Phe Ala Leu Pro Gly Leu Leu His  
 1                                      5                                      10                                      15  
 Ala Gln Gln Ala Cys Ser Arg Gly Ala Cys Tyr Pro Pro Val Gly Asp  
 20                                      25                                      30  
 Leu Leu Val Gly Arg Thr Arg Phe Leu Arg Ala Ser Ser Thr Cys Gly  
 35                                      40                                      45  
 Leu Thr Lys Pro Glu Thr Tyr Cys Thr Gln Tyr Gly Glu Trp Gln Met  
 50                                      55                                      60  
 Lys Cys Cys Lys Cys Asp Ser Arg Gln Pro His Asn Tyr Tyr Ser His  
 65                                      70                                      75                                      80  
 Arg Val Glu Asn Val Ala Ser Ser Ser Gly Pro Met Arg Trp Trp Gln

73

	85	90	95
Ser Gln Asn Asp Val Asn Pro Val Ser Leu Gln Leu Asp Leu Asp Arg			
100	105	110	
Arg Phe Gln Leu Gln Glu Val Met Met Glu Phe Gln Gly Pro Met Pro			
115	120	125	
Ala Gly Met Leu Ile Glu Arg Ser Ser Asp Phe Gly Lys Thr Trp Arg			
130	135	140	
Val Tyr Gln Tyr Leu Ala Ala Asp Cys Thr Ser Thr Phe Pro Arg Val			
145	150	155	160
Arg Gln Gly Arg Pro Gln Ser Trp Gln Asp Val Arg Cys Gln Ser Leu			
165	170	175	
Pro Gln Arg Pro Asn Ala Arg Leu Asn Gly Gly Lys Val Gln Leu Asn			
180	185	190	
Leu Met Asp Leu Val Ser Gly Ile Pro Ala Thr Gln Ser Gln Lys Ile			
195	200	205	
Gln Glu Val Gly Glu Ile Thr Asn Leu Arg Val Asn Phe Thr Arg Leu			
210	215	220	
Ala Pro Val Pro Gln Arg Gly Tyr His Pro Pro Ser Ala Tyr Tyr Ala			
225	230	235	240
Val Ser Gln Leu Arg Leu Gln Gly Ser Cys Phe Cys His Gly His Ala			
245	250	255	
Asp Arg Cys Ala Pro Lys Pro Gly Ala Ser Ala Gly Pro Ser Thr Ala			
260	265	270	
Val Gln Val His Asp Val Cys Val Cys Gln His Asn Thr Ala Gly Pro			
275	280	285	
Asn Cys Glu Arg Cys Ala Pro Phe Tyr Asn Asn Arg Pro Trp Arg Pro			
290	295	300	
Ala Glu Gly Gln Asp Ala His Glu Cys Gln Arg Cys Asp Cys Asn Gly			
305	310	315	320



74

His Ser Glu Thr Cys His Phe Asp Pro Ala Val Phe Ala Ala Ser Gln  
325 330 335

Gly Ala Tyr Gly Gly Val Cys Asp Asn Cys Arg Asp His Thr Glu Gly  
340 345 350

Lys Asn Cys Glu Arg Cys Gln Leu His Tyr Phe Arg Asn Arg Arg Pro  
355 360 365

Gly Ala Ser Ile Gln Glu Thr Cys Ile Ser Cys Glu Cys Asp Pro Asp  
370 375 380

Gly Ala Val Pro Gly Ala Pro Cys Asp Pro Val Thr Gly Gln Cys Val  
385 390 395 400

Cys Lys Glu His Val Gln Gly Glu Arg Cys Asp Leu Cys Lys Pro Gly  
405 410 415

Phe Thr Gly Leu Thr Tyr Ala Asn Pro Gln Gly Cys His Arg Cys Asp  
420 425 430

Cys Asn Ile Leu Gly Ser Arg Arg Asp Met Pro Cys Asp Glu Glu Ser  
435 440 445

Gly Arg Cys Leu Cys Leu Pro Asn Val Val Gly Pro Lys Cys Asp Gln  
450 455 460

Cys Ala Pro Tyr His Trp Lys Leu Ala Ser Gly Gln Gly Cys Glu Pro  
465 470 475 480

Cys Ala Cys Asp Pro His Asn Ser Leu Ser Pro Gln Cys Asn Gln Phe  
485 490 495

Thr Gly Gln Cys Pro Cys Arg Glu Gly Phe Gly Gly Leu Met Cys Ser  
500 505 510

Ala Ala Ala Ile Arg Gln Cys Pro Asp Arg Thr Tyr Gly Asp Val Ala  
515 520 525

Thr Gly Cys Arg Ala Cys Asp Cys Asp Phe Arg Gly Thr Glu Gly Pro  
530 535 540

Gly Cys Asp Lys Ala Ser Gly Arg Cys Leu Cys Arg Pro Gly Leu Thr

75

545	550	555	560
Gly Pro Arg Cys Asp Gln Cys Gln Arg Gly Tyr Cys Asn Arg Tyr Pro			
565	570	575	
Val Cys Val Ala Cys His Pro Cys Phe Gln Thr Tyr Asp Ala Asp Leu			
580	585	590	
Arg Glu Gln Ala Leu Arg Phe Gly Arg Leu Arg Asn Ala Thr Ala Ser			
595	600	605	
Leu Trp Ser Gly Pro Gly Leu Glu Asp Arg Gly Leu Ala Ser Arg Ile			
610	615	620	
Leu Asp Ala Lys Ser Lys Ile Glu Gln Ile Arg Ala Val Leu Ser Ser			
625	630	635	640
Pro Ala Val Thr Glu Gln Glu Val Ala Gln Val Ala Ser Ala Ile Leu			
645	650	655	
Ser Leu Arg Arg Thr Leu Gln Gly Leu Gln Leu Asp Leu Pro Leu Glu			
660	665	670	
Glu Glu Thr Leu Ser Leu Pro Arg Asp Leu Glu Ser Leu Asp Arg Ser			
675	680	685	
Phe Asn Gly Leu Leu Thr Met Tyr Gln Arg Lys Arg Glu Gln Phe Glu			
690	695	700	
Lys Ile Ser Ser Ala Asp Pro Ser Gly Ala Phe Arg Met Leu Ser Thr			
705	710	715	720
Ala Tyr Glu Gln Ser Ala Gln Ala Ala Gln Gln Val Ser Asp Ser Ser			
725	730	735	
Arg Leu Leu Asp Gln Leu Arg Asp Ser Arg Arg Glu Ala Glu Arg Leu			
740	745	750	
Val Arg Gln Ala Gly Gly Gly Gly Thr Gly Ser Pro Lys Leu Val			
755	760	765	
Ala Leu Arg Leu Glu Met Ser Ser Leu Pro Asp Leu Thr Pro Thr Phe			
770	775	780	

76

Asn Lys Leu Cys Gly Asn Ser Arg Gln Met Ala Cys Thr Pro Ile Ser  
 785                      790                      795                      800  
 Cys Pro Gly Glu Leu Cys Pro Gln Asp Asn Gly Thr Ala Cys Gly Ser  
                          805                      810                      815  
 Arg Cys Arg Gly Val Leu Pro Arg Ala Gly Gly Ala Phe Leu Met Ala  
                          820                      825                      830  
 Gly Gln Val Ala Glu Gln Leu Arg Gly Phe Asn Ala Gln Leu Gln Arg  
                          835                      840                      845  
 Thr Arg Gln Met Ile Arg Ala Ala Glu Glu Ser Ala Ser Gln Ile Gln  
                          850                      855                      860  
 Ser Ser Ala Gln Arg Leu Glu Thr Gln Val Ser Ala Ser Arg Ser Gln  
 865                      870                      875                      880  
 Met Glu Glu Asp Val Arg Arg Thr Arg Leu Leu Ile Gln Gln Val Arg  
                          885                      890                      895  
 Asp Phe Leu Thr Asp Pro Asp Thr Asp Ala Ala Thr Ile Gln Glu Val  
                          900                      905                      910  
 Ser Glu Ala Val Leu Ala Leu Trp Leu Pro Thr Asp Ser Ala Thr Val  
                          915                      920                      925  
 Leu Gln Lys Met Asn Glu Ile Gln Ala Ile Ala Ala Arg Leu Pro Asn  
                          930                      935                      940  
 Val Asp Leu Val Leu Ser Gln Thr Lys Gln Asp Ile Ala Arg Ala Arg  
 945                      950                      955                      960  
 Arg Leu Gln Ala Glu Ala Glu Glu Ala Arg Ser Arg Ala His Ala Val  
                          965                      970                      975  
 Glu Gly Gln Val Glu Asp Val Val Gly Asn Leu Arg Gln Gly Thr Val  
                          980                      985                      990  
 Ala Leu Gln Glu Ala Gln Asp Thr Met Gln Gly Thr Ser Arg Ser Leu  
                          995                      1000                      1005  
 Arg Leu Ile Gln Asp Arg Val Ala Glu Val Gln Gln Val Leu Arg Pro

1010 1015 1020  
Ala Glu Lys Leu Val Thr Ser Met Thr Lys Gln Leu Gly Asp Phe Trp  
1025 1030 1035 1040  
Thr Arg Met Glu Glu Leu Arg His Gln Ala Arg Gln Gln Gly Ala Glu  
1045 1050 1055  
Ala Val Gln Ala Gln Gln Leu Ala Glu Gly Ala Ser Glu Gln Ala Leu  
1060 1065 1070  
Ser Ala Gln Glu Gly Phe Glu Arg Ile Lys Gln Lys Tyr Ala Glu Leu  
1075 1080 1085  
Lys Asp Arg Leu Gly Gln Ser Ser Met Leu Gly Glu Gln Gly Ala Arg  
1090 1095 1100  
Ile Gln Ser Val Lys Thr Glu Ala Glu Glu Leu Phe Gly Glu Thr Met  
1105 1110 1115 1120  
Glu Met Met Asp Arg Met Lys Asp Met Glu Leu Glu Leu Leu Arg Gly  
1125 1130 1135  
Ser Gln Ala Ile Met Leu Arg Ser Ala Asp Leu Thr Gly Leu Glu Lys  
1140 1145 1150  
Arg Val Glu Gln Ile Arg Asp His Ile Asn Gly Arg Val Leu Tyr Tyr  
1155 1160 1165  
Ala Thr Cys Lys  
1170

Sequence No.: 8

Sequence length: 122

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

78

Organism species: *Homo sapiens*

Cell kind: Stomach cancer

Clone name: HP10298

Sequence description

Met Gly Leu Leu Leu Leu Val Pro Leu Leu Leu Leu Pro Gly Ser Tyr

1 5 10 15

Gly Leu Pro Phe Tyr Asn Gly Phe Tyr Tyr Ser Asn Ser Ala Asn Asp

20 25 30

Gln Asn Leu Gly Asn Gly His Gly Lys Asp Leu Leu Asn Gly Val Lys

35 40 45

Leu Val Val Glu Thr Pro Glu Glu Thr Leu Phe Thr Arg Ile Leu Thr

50 55 60

Val Gly Pro Gln Ser Leu Gly Ser Glu Ala Leu Ala Ser Pro Thr Arg

65 70 75 80

Arg Ala Ala Cys Thr Val Phe Thr Ala Thr Ala Ser Thr Arg Thr Trp

85 90 95

Gly Pro Pro Leu Pro His Ser Leu Thr Gly Cys Val Phe Ile Glu Trp

100 105 110

Phe Val Phe Pro Cys Gly Leu Glu Pro Phe

115 120

Sequence No.: 9

Sequence length: 175

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

79

Cell kind: Stomach cancer

Clone name: HP10368

## Sequence description

Met Glu Lys Ile Pro Val Ser Ala Phe Leu Leu Leu Val Ala Leu Ser  
1 5 10 15  
Tyr Thr Leu Ala Arg Asp Thr Thr Val Lys Pro Gly Ala Lys Lys Asp  
20 25 30  
Thr Lys Asp Ser Arg Pro Lys Leu Pro Gln Thr Leu Ser Arg Gly Trp  
35 40 45  
Gly Asp Gln Leu Ile Trp Thr Gln Thr Tyr Glu Glu Ala Leu Tyr Lys  
50 55 60  
Ser Lys Thr Ser Asn Lys Pro Leu Met Ile Ile His His Leu Asp Glu  
65 70 75 80  
Cys Pro His Ser Gln Ala Leu Lys Lys Val Phe Ala Glu Asn Lys Glu  
85 90 95  
Ile Gln Lys Leu Ala Glu Gln Phe Val Leu Leu Asn Leu Val Tyr Glu  
100 105 110  
Thr Thr Asp Lys His Leu Ser Pro Asp Gly Gln Tyr Val Pro Arg Ile  
115 120 125  
Met Phe Val Asp Pro Ser Leu Thr Val Arg Ala Asp Ile Thr Gly Arg  
130 135 140  
Tyr Ser Asn Arg Leu Tyr Ala Tyr Glu Pro Ala Asp Thr Ala Leu Leu  
145 150 155 160  
Leu Asp Asn Met Lys Lys Ala Leu Lys Leu Leu Lys Thr Glu Leu  
165 170 175

Sequence No.: 10

Sequence length: 462

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Fibrosarcoma

Cell line: HT-1080

Clone name: HP00658

Sequence description

ATGAAGGTCT CCGCGGCAGC CCTCGCTGTC ATCCTCATTG CTA CTGCGCTCCT	60
GCATCTGCCT CCCCATATTC CTCGGACACC ACACCCTGCT GCTTTGCCTA CATTGCCCCG	120
CCACTGCCCC GTGCCCACAT CAAGGAGTAT TTCTACACCA GTGGCAAGTG CTCCAACCCA	180
GCAGTCGTCC ACAGGTCAAG GATGCCAAAG AGAGAGGGAC AGCAAGTCTG GCAGGATTTC	240
CTGTATGACT CCCGGCTGAA CAAGGGCAAG CTTTGTCAAC CGAAAGAACC GCCAAGTGTG	300
TGCCAACCCA GAGAAGAAAT GGGTTCGGGA GTACATCAAC TCTTTGGAGA TGAGCTAGGA	360
TGGAGAGTCC TTGAACCTGA ACTTACACAA ATTTGCCTGT TTCTGCTTGC TCTTGTCTA	420
GCTTGGGAGG CTTCCCTCA CTATCCTACC CCACCCGCTC CT	462

Sequence No.: 11

Sequence length: 945

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Epidermoid carcinoma

Cell line: KB

Clone name: HP00714

Sequence description

ATGGACCTGC GACAGTTTCT TATGTGCCTG TCCCTGTGCA CAGCCTTTGC CTTGAGCAAA	60
CCCACAGAAA AGAAGGACCG TGTACATCAT GAGCCTCAGC TCAGTGACAA GGTTACAAT	120
GATGCTCAGA GTTTTGATTA TGACCATGAT GCCTTCTTGG GTGCTGAAGA AGCAAAGACC	180
TTTGATCAGC TGACACCAGA AGAGAGCAAG GAAAGGCTTG GAAAGATTGT AAGTAAAATA	240
GATGGCGACA AGGACGGGTT TGTCACTGTG GATGAGCTCA AAGACTGGAT TAAATTTGCA	300
CAAAAGCGCT GGATTTACGA GGATGTAGAG CGACAGTGGG AGGGGCATGA CCTCAATGAG	360
GACGGCCTCG TTTCCTGGGA GGAGTATAAA AATGCCACCT ACGGCTACGT TTTAGATGAT	420
CCAGATCCTG ATGATGGATT TAACTATAAA CAGATGATGG TTAGAGATGA GCGGAGGTTT	480
AAAATGGCAG ACAAGGATGG AGACCTCATT GCCACCAAGG AGGAGTTCAC AGCTTTCCTG	540
CACCTGAGG AGTATGACTA CATGAAAGAT ATAGTAGTAC AGGAAACAAT GGAAGATATA	600
GATAAGAATG CTGATGGTTT CATTGATCTA GAAGAGTATA TTGGTGACAT GTACAGCCAT	660
GATGGGAATA CTGATGAGCC AGAATGGGTA AAGACAGAGC GAGAGCAGTT TGTTGAGTTT	720
CGGGATAAGA ACCGTGATGG GAAGATGGAC AAGGAAGAGA CCAAAGACTG GATCCTTCCC	780
TCAGACTATG ATCATGCAGA GGCAGAAGCC AGGCACCTGG TCTATGAATC AGACCAAAAC	840
AAGGATGGCA AGCTTACCAA GGAGGAGATC GTTGACAAGT ATGACTTATT TGTGGCAGC	900
CAGGCCACAG ATTTTGGGGA GGCCTTAGTA CGGCATGATG AGTTC	945

Sequence No.: 12

Sequence length: 474

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Stomach cancer

Clone name: HP00876



## Sequence description

ATGGCTTCCA GAAGCATGCG GCTGCTCCTA TTGCTGAGCT GCCTGGCCAA AACAGGAGTC	60
CTGGGTGATA TCATCATGAG ACCCAGCTGT GCTCCTGGAT GGTTTACCA CAAGTCCAAT	120
TGCTATGGTT ACTTCAGGAA GCTGAGGAAC TGGTCTGATG CCGAGCTCGA GTGTCAGTCT	180
TACGGAAACG GAGCCCACCT GGCATCTATC CTGAGTTTAA AGGAAGCCAG CACCATAGCA	240
GAGTACATAA GTGGCTATCA GAGAAGCCAG CCGATATGGA TTGGCCTGCA CGACCCACAG	300
AAGAGGCAGC AGTGGCAGTG GATTGATGGG GCCATGTATC TGTACAGATC CTGGTCTGGC	360
AAGTCCATGG GTGGGAACAA GCACTGTGCT GAGATGAGCT CCAATAACAA CTTTTTAACT	420
TGGAGCAGCA ACGAATGCAA CAAGCGCCAA CACTTCCTGT GCAAGTACCG ACCA	474

Sequence No.: 13

Sequence length: 1128

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Liver

Clone name: HP01134

## Sequence description

ATGGTTTGA AAGTAGCTGT ATTCCTCAGT GTGGCCCTGG GCATTGGTGC CGTTCCTATA	60
GATGATCCTG AAGATGGAGG CAAGCACTGG GTGGTGATCG TGGCAGGTTT AAATGGCTGG	120
TATAATTATA GGCACCAGGC AGACGCGTGC CATGCCCTACC AGATCATTCA CCGCAATGGG	180
ATTCCTGACG AACAGATCGT TGTGATGATG TACGATGACA TTGCTTACTC TGAAGACAAT	240
CCCACTCCAG GAATTGTGAT CAACAGGCCA AATGGCACAG ATGTCTATCA GGGAGTCCCG	300
AAGGACTACA CTGGAGAGGA TGTTACCCCA CAAAATTTCC TTGCTGTGTT GAGAGGCGAT	360
GCAGAAGCAG TGAAGGGCAT AGGATCCGGC AAAGTCCTGA AGAGTGGCCC CCAGGATCAC	420
GTGTTCAATT ACTTCACTGA CCATGGATCT ACTGGAATAC TGGTTTTTCC CAATGAAGAT	480

CTTCATGTAA AGGACCTGAA TGAGACCATC CATTACATGT ACAAACACAA AATGTACCGA	540
AAGATGGTGT TCTACATTGA AGCCTGTGAG TCTGGGTCCA TGATGAACCA CCTGCCGGAT	600
AACATCAATG TTTATGCAAC TACTGCTGCC AACCCAGAG AGTCGTCCTA CGCCTGTTAC	660
TATGATGAGA AGAGGTCCAC GTACCTGGGG GACTGGTACA GCGTCAACTG GATGGAAGAC	720
TCGGACGTGG AAGATCTGAC TAAAGAGACC CTGCACAAGC AGTACCACCT GGTAATTCG	780
CACACCAACA CCAGCCACGT CATGCAGTAT GGAAACAAAA CAATCTCCAC CATGAAAGTG	840
ATGCAGTTTC AGGGTATGAA ACGCAAAGCC AGTTCTCCCG TCCCCCTACC TCCAGTCACA	900
CACCTTGACC TCACCCCCAG CCCTGATGTG CCTCTACCA TCATGAAAAG GAAACTGATG	960
AACACCAATG ATCTGGAGGA GTCCAGGCAG CTCACGGAGG AGATCCAGCG GCATCTGGAT	1020
TACGAGTATG CGTTGAGACA TTTGTACGTG CTGGTCAACC TTTGTGAGAA GCCGTATCCG	1080
CTTCACAGGA TAAAATTGTC CATGGACCAC GTGTGCCTTG GTCACTAC	1128

Sequence No.: 14

Sequence length: 519

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Epidermoid carcinoma

Cell line: KB

Clone name: HP10029

Sequence description

ATGGCGGCGC CCAGCGGAGG GTGGAACGGC GTCCGCGCGA GCTTGTGGGC CGCGCTGCTC	60
CTAGGGGCGG TGGCGCTGAG GCCGGCGGAG GCGGTGTCCG AGCCCACGAC CGTGGCGTTT	120
GACGTGCGGC CCGCGGGCGT CGTGCAATCC TTCTCCATA ACGTGGGCCC GGGGGACAAA	180
TATACGTGTA TGTTCACTTA CGCCTCTCAA GGAGGGACCA ATGAGCAATG GCAGATGAGT	240
CTGGGGACCA GCGAAGACCA CCAGCACTTC ACCTGCACCA TCTGGAGGCC CCAGGGGAAG	300

84

TCCTATCTGT ACTTCACACA GTTCAAGGCA GAGGTGCGGG GCGCTGAGAT TGAGTACGCC 360  
ATGGCCTACT CTAAAGCCGC ATTTGAAAGG GAAAGTGATG TCCCTCTGAA AACTGAGGAA 420  
TTTGAAGTGA CCAAAACAGC AGTGGCTCAC AGGCCCGGGG CATTCAAAGC TGAGCTGTCC 480  
AAGCTGGTGA TTGTGGCCAA GGCATCGCGC ACTGAGCTG 519

Sequence No.: 15

Sequence length: 219

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Epidermoid carcinoma

Cell line: KB

Clone name: HP10189

Sequence description

ATGGGGGTGA AGCTGGAGAT ATTTGGGATG ATAATCTACC TCACTTTCCC TGTGGCTATG 60  
TTCTGGGTTT CCAATCAGGC CGAGTGGTTT GAGGACGATG TCATACAGCG CAAGAGGGAG 120  
CTGTGGCCAC CTGAGAAGCT TCAAGAGATA GAGGAATTCA AAGAGAGGTT ACGGAAGCGG 180  
CGGGAGGAGA AGCTCCTTCG CGACGCCCAG CAGAACTCC 219

Sequence No.: 16

Sequence length: 3516

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Lymphoma

Cell line: U937

Clone name: HP10269

Sequence description

ATGAGACCAT TCTTCCTCTT GTGTTTGGCC CTGCCTGGCC TCCTGCATGC CCAACAAGCC	60
TGCTCCCGTG GGGCTGCTA TCCACCTGTT GGGGACCTGC TTGTTGGGAG GACCCGGTTT	120
CTCCGAGCTT CATCTACCTG TGGACTGACC AAGCCTGAGA CCTACTGCAC CCAGTATGGC	180
GAGTGGCAGA TGAAATGCTG CAAGTGTGAC TCCAGGCAGC CTCACAACTA CTACAGTCAC	240
CGAGTAGAGA ATGTGGCTTC ATCCTCCGGC CCCATGCGCT GGTGGCAGTC CCAGAATGAT	300
GTGAACCCTG TCTCTCTGCA GCTGGACCTG GACAGGAGAT TCCAGCTTCA AGAAGTCATG	360
ATGGAGTTCC AGGGGCCCAT GCCTGCCGGC ATGCTGATTG AGCGCTCCTC AGACTTCGGT	420
AAGACCTGGC GAGTGTACCA GTACCTGGCT GCCGACTGCA CCTCCACCTT CCCTCGGGTC	480
CGCCAGGGTC GGCCTCAGAG CTGGCAGGAT GTTCGGTGCC AGTCCCTGCC TCAGAGGCCT	540
AATGCACGCC TAAATGGGGC GAAGGTCCAA CTTAACCTTA TGGATTTAGT GTCTGGGATT	600
CCAGCAACTC AAAGTCAAAA AATTCAAGAG GTGGGGGAGA TCACAAACTT GAGAGTCAAT	660
TTCACCAGGC TGGCCCTGT GCCCCAAAGG GGCTACCACC CTCCCAGCGC CTACTATGCT	720
GTGTCCCAGC TCCGTCTGCA GGGGAGCTGC TTCTGTACAG GCCATGCTGA TCGCTGCGCA	780
CCCAAGCCTG GGGCCTCTGC AGGCCCTCC ACCGCTGTGC AGGTCCACGA TGTCTGTGTC	840
TGCCAGCACA ACACTGCCGG CCCAAATTGT GAGCGCTGTG CACCCTTCTA CAACAACCGG	900
CCCTGGAGAC CGGCGGAGGG CCAGGACGCC CATGAATGCC AAAGTGCGA CTGCAATGGG	960
CACTCAGAGA CATGTCACTT TGACCCCGCT GTGTTTGCCG CCAGCCAGGG GGCATATGGA	1020
GGTGTGTGTG ACAATTGCCG GGACCACACC GAAGGCAAGA ACTGTGAGCG GTGTCAGCTG	1080
CACTATTTCC GGAACCGGCG CCCGGGAGCT TCCATTGAGG AGACCTGCAT CTCCTGCGAG	1140
TGTGATCCGG ATGGGGCAGT GCCAGGGGCT CCCTGTGACC CAGTGACCGG GCAGTGTGTG	1200
TGCAAGGAGC ATGTGCAGGG AGAGCGCTGT GACCTATGCA AGCCGGGCTT CACTGGACTC	1260
ACCTACGCCA ACCCGCAGGG CTGCCACCGC TGTGACTGCA ACATCCTGGG GTCCCGGAGG	1320
GACATGCCGT GTGACGAGGA GAGTGGGCGC TGCCTTTGTC TGCCCAACGT GGTGGGTCCC	1380
AAATGTGACC AGTGTGCTCC CTACCACTGG AAGCTGGCCA GTGGCCAGGG CTGTGAACCG	1440

TGTGCTGCG	ACCCGCACAA	CTCCCTCAGC	CCACAGTGCA	ACCAGTTCAC	AGGGCAGTGC	1500
CCCTGTGCGG	AAGGCTTTGG	TGGCCTGATG	TGCAGCGCTG	CAGCCATCCG	CCAGTGTCCA	1560
GACCGGACCT	ATGGAGACGT	GGCCACAGGA	TGCCGAGCCT	GTGACTGTGA	TTTCCGGGGA	1620
ACAGAGGGCC	CGGGCTGCCA	CAAGGCATCA	GGCCGCTGCC	TCTGCCGCCC	TGGCTTGACC	1680
GGGCCCCGCT	GTGACCAGTG	CCAGCGAGGC	TACTGCAATC	GCTACCCGGT	GTGCGTGGCC	1740
TGCCACCCTT	GCTTCCAGAC	CTATGATGCG	GACCTCCGGG	AGCAGGCCCT	GCGCTTTGGT	1800
AGACTCCGCA	ATGCCACCGC	CAGCCTGTGG	TCAGGGCCTG	GGCTGGAGGA	CCGTGGCCTG	1860
GCCTCCCGGA	TCCTAGATGC	AAAGAGTAAG	ATTGAGCAGA	TCCGAGCAGT	TCTCAGCAGC	1920
CCCGCAGTCA	CAGAGCAGGA	GGTGGCTCAG	GTGGCCAGTG	CCATCCTCTC	CCTCAGGCGA	1980
ACTCTCCAGG	GCCTGCAGCT	GGATCTGCCC	CTGGAGGAGG	AGACGTTGTC	CCTTCCGAGA	2040
GACCTGGAGA	GTCTTGACAG	AAGCTTCAAT	GGTCTCCTTA	CTATGTATCA	GAGGAAGAGG	2100
GAGCAGTTTG	AAAAAATAAG	CAGTGCTGAT	CCTTCAGGAG	CCTTCCGGAT	GCTGAGCACA	2160
GCCTACGAGC	AGTCAGCCCA	GGCTGCTCAG	CAGGTCTCCG	ACAGCTCGCG	CCTTTTGGAC	2220
CAGCTCAGGG	ACAGCCGGAG	AGAGGCAGAG	AGGCTGGTGC	GGCAGGCGGG	AGGAGGAGGA	2280
GGCACC GGCA	GCCCCAAGCT	TGTGGCCCTG	AGGCTGGAGA	TGTCTTCGTT	GCCTGACCTG	2340
ACACCCACCT	TCAACAAGCT	CTGTGGCAAC	TCCAGGCAGA	TGGCTTGAC	CCCAATATCA	2400
TGCCCTGGTG	AGCTATGTCC	CCAAGACAAT	GGCACAGCCT	GTGGCTCCCG	CTGCAGGGGT	2460
GTCTTTCCCA	GGGCCGGTGG	GGCCTTCTTG	ATGGCGGGGC	AGGTGGCTGA	GCAGCTGCGG	2520
GGCTTCAATG	CCCAGCTCCA	GCGGACCAGG	CAGATGATTA	GGGCAGCCGA	GGAATCTGCC	2580
TCACAGATTG	AATCCAGTGC	CCAGCGCTTG	GAGACCCAGG	TGAGCGCCAG	CCGCTCCAG	2640
ATGGAGGAAG	ATGTCAGACG	CACACGGCTC	CTAATCCAGC	AGGTCCGGGA	CTTCCTAACA	2700
GACCCCGACA	CTGATGCAGC	CACTATCCAG	GAGGTCAGCG	AGGCCGTGCT	GGCCCTGTGG	2760
CTGCCCACAG	ACTCAGCTAC	TGTTCTGCAG	AAGATGAATG	AGATCCAGGC	CATTGCAGCC	2820
AGGCTCCCCA	ACGTGGACTT	GGTGCTGTCC	CAGACCAAGC	AGGACATTGC	GCGTGCCCGC	2880
CGGTTGCAGG	CTGAGGCTGA	GGAAGCCAGG	AGCCGAGCCC	ATGCAGTGA	GGGCCAGGTG	2940
GAAGATGTGG	TTGGGAACCT	GCGGCAGGGG	ACAGTGGCAC	TGCAGGAAGC	TCAGGACACC	3000
ATGCAAGGCA	CCAGCCGCTC	CCTTCGGCTT	ATCCAGGACA	GGGTTGCTGA	GGTTCAGCAG	3060
GTA CTGCGG	CAGCAGAAAA	GCTGGTGACA	AGCATGACCA	AGCAGCTGGG	TGACTTCTGG	3120
ACACGGATGG	AGGAGCTCCG	CCACCAAGCC	GCGCAGCAGG	GGCAGAGGC	AGTCCAGGCC	3180

87

CAGCAGCTTG CGGAAGGTGC CAGCGAGCAG GCATTGAGTG CCCAAGAGGG ATTTGAGAGA 3240  
ATAAAACAAA AGTATGCTGA GTTGAAGGAC CGGTTGGGTC AGAGTTCCAT GCTGGGTGAG 3300  
CAGGGTGCCC GGATCCAGAG TGTGAAGACA GAGGCAGAGG AGCTGTTTGG GGAGACCATG 3360  
GAGATGATGG ACAGGATGAA AGACATGGAG TTGGAGCTGC TCGGGGGCAG CCAGGCCATC 3420  
ATGCTGCGCT CAGCGGACCT GACAGGACTG GAGAAGCGTG TGGAGCAGAT CCGTGACCAC 3480  
ATCAATGGGC GCGTGCTCTA CTATGCCACC TGCAAG 3516

Sequence No.: 17

Sequence length: 366

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Stomach cancer

Clone name: HP10298

Sequence description

ATGGGCCTGT TGCTCCTGGT CCCATTGCTC CTGCTGCCCG GCTCCTACGG ACTGCCCTTC 60  
TACAACGGCT TCTACTACTC CAACAGCGCC AACGACCAGA ACCTAGGCAA CGGTCATGGC 120  
AAAGACCTCC TTAATGGAGT GAAGCTGGTG GTGGAGACAC CCGAGGAGAC CCTGTTCACC 180  
CGCATCTAA CTGTGGGCCC CCAGAGCCTG GGGTCCGAAG CTTTGGCTTC CCCGACCCGC 240  
AGAGCCGCTT GTACGGTGTT TACTGCTACC GCCAGCACTA GGACCTGGGG CCCTCCCCTG 300  
CCGCATTCCC TCACTGGCTG TGTATTTATT GAGTGGTTCG TTTTCCCTTG TGGGTGGAG 360  
CCATT 366

Sequence No.: 18

Sequence length: 525

Sequence type: Nucleic acid

88

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Stomach cancer

Clone name: HP10368

Sequence description

ATGGAGAAAA TTCCAGTGTG AGCATTCTTG CTCCTTGTGG CCCTCTCCTA CACTCTGGCC	60
AGAGATACCA CAGTCAAACC TGGAGCCAAA AAGGACACAA AGGACTCTCG ACCCAAAC TG	120
CCCCAGACCC TCTCCAGAGG TTGGGGTGAC CAACTCATCT GGA CT CAGAC ATATGAAGAA	180
GCTCTATATA AATCCAAGAC AAGCAACAAA CCCTTGATGA TTATTCATCA CTTGGATGAG	240
TGCCCACACA GTCAAGCTTT AAAGAAAGTG TTTGCTGAAA ATAAAGAAAT CCAGAAATTG	300
GCAGAGCAGT TTGTCCTCCT CAATCTGGTT TATGAAACAA CTGACAAACA CCTTTCTCCT	360
GATGGCCAGT ATGTCCCCAG GATTATGTTT GTTGACCCAT CTCTGACAGT TAGAGCCGAT	420
ATCACTGGAA GATATTCAAA CCGTCTCTAT GCTTACGAAC CTGCAGATAC AGCTCTGTTG	480
CTTGACAACA TGAAGAAAGC TCTCAAGTTG CTGAAGACTG AATTG	525

Sequence No.: 19

Sequence length: 1296

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Fibrosarcoma

Cell line: HT-1080

Clone name: HP00658

89

## Sequence characteristics:

Code representing characteristics: CDS

Existence site: 56.. 520

Characterization method: E

## Sequence description

CCTGCAGAGG ATCAAGACAG CACGTGGACC TCGCACAGCC TCTCCCACAG GTACC ATG	58
	Met
	1
AAG GTC TCC GCG GCA GCC CTC GCT GTC ATC CTC ATT GCT ACT GCC CTC	106
Lys Val Ser Ala Ala Ala Leu Ala Val Ile Leu Ile Ala Thr Ala Leu	
5 10 15	
TGC GCT CCT GCA TCT GCC TCC CCA TAT TCC TCG GAC ACC ACA CCC TGC	154
Cys Ala Pro Ala Ser Ala Ser Pro Tyr Ser Ser Asp Thr Thr Pro Cys	
20 25 30	
TGC TTT GCC TAC ATT GCC CGC CCA CTG CCC CGT GCC CAC ATC AAG GAG	202
Cys Phe Ala Tyr Ile Ala Arg Pro Leu Pro Arg Ala His Ile Lys Glu	
35 40 45	
TAT TTC TAC ACC AGT GGC AAG TGC TCC AAC CCA GCA GTC GTC CAC AGG	250
Tyr Phe Tyr Thr Ser Gly Lys Cys Ser Asn Pro Ala Val Val His Arg	
50 55 60 65	
TCA AGG ATG CCA AAG AGA GAG GGA CAG CAA GTC TGG CAG GAT TTC CTG	298
Ser Arg Met Pro Lys Arg Glu Gly Gln Gln Val Trp Gln Asp Phe Leu	
70 75 80	
TAT GAC TCC CGG CTG AAC AAG GGC AAG CTT TGT CAC CCG AAA GAA CCG	346
Tyr Asp Ser Arg Leu Asn Lys Gly Lys Leu Cys His Pro Lys Glu Pro	
85 90 95	
CCA AGT GTG TGC CAA CCC AGA GAA GAA ATG GGT TCG GGA GTA CAT CAA	394
Pro Ser Val Cys Gln Pro Arg Glu Glu Met Gly Ser Gly Val His Gln	
100 105 110	



90

CTC TTT GGA GAT GAG CTA GGA TGG AGA GTC CTT GAA CCT GAA CTT ACA 442  
Leu Phe Gly Asp Glu Leu Gly Trp Arg Val Leu Glu Pro Glu Leu Thr  
115 120 125  
CAA ATT TGC CTG TTT CTG CTT GCT CTT GTC CTA GCT TGG GAG GCT TCC 490  
Gln Ile Cys Leu Phe Leu Leu Ala Leu Val Leu Ala Trp Glu Ala Ser  
130 135 140 145  
CCT CAC TAT CCT ACC CCA CCC GCT CCT TGAAGGGCCC AGA 530  
Pro His Tyr Pro Thr Pro Pro Ala Pro  
150  
TTCTACCACA CAGCAGCAGT TACAAAAACC TTCCCCAGGC TGGACGTGGT GGCTCACGCC 590  
TGTAATCCCA GCACTTTGGG AGGCCAAGGT GGGTGGATCA CTTGAGGTCA GGAGTTCGAG 650  
ACCAGCCTGG CCAACATGAT GAAACCCCAT CTCTACTAAA AATACAAAAA ATTAGCCGGG 710  
CGTGGTAGCG GCGCCTGTA GTCCCAGCTA CTCGGGAGGC TGAGGCAGGA GAATGGCGTG 770  
AACCCGGGAG GCGGAGCTTG CAGTGAGCCG AGATCGCGCC ACTGCACTCC AGCCTGGGCG 830  
ACAGAGCGAG ACTCCGTCTC AAAAAAAAAA AAAAAAAAAA AAATACAAAA ATTAGCCGGG 890  
CGTGGTGGCC CACGCCTGTA ATCCCAGCTA CTCGGGAGGC TAAGGCAGGA AAATTGTTTG 950  
AACCCAGGAG GTGGAGGCTG CAGTGAGCTG AGATTGTGCC ACTTCACTCC AGCCTGGGTG 1010  
ACAAAGTGAG ACTCCGTAC AACAACAACA AAAAAAGCT TCCCCAACTA AAGCCTAGAA 1070  
GAGCTTCTGA GCGCCTGCTT TGTCAAAAGG AAGTCTCTAG GTTCTGAGCT CTGGCTTTGC 1130  
CTTGGCTTTG CCAGGGCTCT GTGACCAGGA AGGAAGTCAG CATGCCTCTA GAGGCAAGGA 1190  
GGGGAGGAAC GCTGCACTCT TAAGCTTCGG CCGTCTCAAC CCCTCACAGG AGCTTACTGG 1250  
CAAACATGAA AAATCGGCTT ACCATTAAAG TTCTCAATGC AACCAT 1296

Sequence No.: 20

Sequence length: 3311

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

## Original source:

Organism species: *Homo sapiens*  
 Cell kind: Epidermoid carcinoma  
 Cell line: KB  
 Clone name: HP00714

## Sequence characteristics:

Code representing characteristics: CDS  
 Existence site: 57.. 1004  
 Characterization method: E

## Sequence description

```

GAGCGGCGGC CACGGCATCC TGTGCTGTGG GGGCTACGAG GAAAGATCTA ATTATC ATG      59
                                                    Met
                                                    1

GAC CTG CGA CAG TTT CTT ATG TGC CTG TCC CTG TGC ACA GCC TTT GCC      107
Asp Leu Arg Gln Phe Leu Met Cys Leu Ser Leu Cys Thr Ala Phe Ala

          5              10              15

TTG AGC AAA CCC ACA GAA AAG AAG GAC CGT GTA CAT CAT GAG CCT CAG      155
Leu Ser Lys Pro Thr Glu Lys Lys Asp Arg Val His His Glu Pro Gln

          20              25              30

CTC AGT GAC AAG GTT CAC AAT GAT GCT CAG AGT TTT GAT TAT GAC CAT      203
Leu Ser Asp Lys Val His Asn Asp Ala Gln Ser Phe Asp Tyr Asp His

          35              40              45

GAT GCC TTC TTG GGT GCT GAA GAA GCA AAG ACC TTT GAT CAG CTG ACA      251
Asp Ala Phe Leu Gly Ala Glu Glu Ala Lys Thr Phe Asp Gln Leu Thr

          50              55              60              65

CCA GAA GAG AGC AAG GAA AGG CTT GGA AAG ATT GTA AGT AAA ATA GAT      299
Pro Glu Glu Ser Lys Glu Arg Leu Gly Lys Ile Val Ser Lys Ile Asp

          70              75              80

GGC GAC AAG GAC GGG TTT GTC ACT GTG GAT GAG CTC AAA GAC TGG ATT      347

```

92

Gly Asp Lys Asp Gly Phe Val Thr Val Asp Glu Leu Lys Asp Trp Ile  
 85 90 95  
 AAA TTT GCA CAA AAG CGC TGG ATT TAC GAG GAT GTA GAG CGA CAG TGG 395  
 Lys Phe Ala Gln Lys Arg Trp Ile Tyr Glu Asp Val Glu Arg Gln Trp  
 100 105 110  
 AAG GGG CAT GAC CTC AAT GAG GAC GGC CTC GTT TCC TGG GAG GAG TAT 443  
 Lys Gly His Asp Leu Asn Glu Asp Gly Leu Val Ser Trp Glu Glu Tyr  
 115 120 125  
 AAA AAT GCC ACC TAC GGC TAC GTT TTA GAT GAT CCA GAT CCT GAT GAT 491  
 Lys Asn Ala Thr Tyr Gly Tyr Val Leu Asp Asp Pro Asp Pro Asp Asp  
 130 135 140 145  
 GGA TTT AAC TAT AAA CAG ATG ATG GTT AGA GAT GAG CGG AGG TTT AAA 539  
 Gly Phe Asn Tyr Lys Gln Met Met Val Arg Asp Glu Arg Arg Phe Lys  
 150 155 160  
 ATG GCA GAC AAG GAT GGA GAC CTC ATT GCC ACC AAG GAG GAG TTC ACA 587  
 Met Ala Asp Lys Asp Gly Asp Leu Ile Ala Thr Lys Glu Glu Phe Thr  
 165 170 175  
 GCT TTC CTG CAC CCT GAG GAG TAT GAC TAC ATG AAA GAT ATA GTA GTA 635  
 Ala Phe Leu His Pro Glu Glu Tyr Asp Tyr Met Lys Asp Ile Val Val  
 180 185 190  
 CAG GAA ACA ATG GAA GAT ATA GAT AAG AAT GCT GAT GGT TTC ATT GAT 683  
 Gln Glu Thr Met Glu Asp Ile Asp Lys Asn Ala Asp Gly Phe Ile Asp  
 195 200 205  
 CTA GAA GAG TAT ATT GGT GAC ATG TAC AGC CAT GAT GGG AAT ACT GAT 731  
 Leu Glu Glu Tyr Ile Gly Asp Met Tyr Ser His Asp Gly Asn Thr Asp  
 210 215 220 225  
 GAG CCA GAA TGG GTA AAG ACA GAG CGA GAG CAG TTT GTT GAG TTT CGG 779  
 Glu Pro Glu Trp Val Lys Thr Glu Arg Glu Gln Phe Val Glu Phe Arg  
 230 235 240

93

GAT AAG AAC CGT GAT GGG AAG ATG GAC AAG GAA GAG ACC AAA GAC TGG	827
Asp Lys Asn Arg Asp Gly Lys Met Asp Lys Glu Glu Thr Lys Asp Trp	
245 250 255	
ATC CTT CCC TCA GAC TAT GAT CAT GCA GAG GCA GAA GCC AGG CAC CTG	875
Ile Leu Pro Ser Asp Tyr Asp His Ala Glu Ala Glu Ala Arg His Leu	
260 265 270	
GTC TAT GAA TCA GAC CAA AAC AAG GAT GGC AAG CTT ACC AAG GAG GAG	923
Val Tyr Glu Ser Asp Gln Asn Lys Asp Gly Lys Leu Thr Lys Glu Glu	
275 280 285	
ATC GTT GAC AAG TAT GAC TTA TTT GTT GGC AGC CAG GCC ACA GAT TTT	971
Ile Val Asp Lys Tyr Asp Leu Phe Val Gly Ser Gln Ala Thr Asp Phe	
290 295 300 305	
GGG GAG GCC TTA GTA CGG CAT GAT GAG TTC TGAGCTACGG AGGAACCCT	1020
Gly Glu Ala Leu Val Arg His Asp Glu Phe	
310 315	
CATTTCTCA AAAGTAATTT ATTTTACAG CTTCTGTTT CACATGAAAT TGTTTGGCT	1080
ACTGAGACTG TTAACATAAA CTTTTAAGA CATGAAAAGG CGTAATGAAA ACCATCCCGT	1140
CCCCATTCCT CCTCCTCTCT GAGGGACTGG AGGGAAGCCG TGCTTCTGAG GAACAACTCT	1200
AATTAGTACA CTTGTGTTTG TAGATTTACA CTTTGTATTA TGTATTAACA TGGCGTGTTT	1260
ATTTTGTAT TTTTCTCTGG TTGGGAGTAT GATATGAAGG ATCAAGATCC TCAACTCACA	1320
CATGTAGACA AACATTAGCT CTTTACTCTT TCTCAACCCC TTTTATGATT TTAATAATTC	1380
TCACTTAACT AATTTTGTA GCGTGAGATC AATAAGAAAT GTTCAGGAGA GAGGAAAGAA	1440
AAAAAATATA TGCTCCACAA TTTATATTTA GAGAGAGAAC ACTTAGTCTT GCCTGTCAAA	1500
AAGTCCAACA TTTCATAGGT AGTAGGGGCC ACATATTACA TTCAGTTGCT ATAGGTCCAG	1560
CAACTGAACC TGCCATTACC TGGGCAAGGA AAGATCCCTT TGCTCTAGGA AAGCTTGGCC	1620
CAAATTGATT TTCTTCTTTT TCCCCCTGTA GGAAGTACTG TTGGCTAATT TTGTCAAGCA	1680
CAGCTGTGGT GGAAGAGTT AGGGCCAGTG TCTTGAAAAT CAATCAAGTA GTGAATGTGA	1740
TCTCTTTGCA GAGCTATAGA TAGAAACAGC TGGAAAATA AAGGAAAAAT ACAAGTGTTT	1800
TCGGGGCATA CATTTTTTTT CTGGGTGTGC ATCTGTTGAA ATGCTCAAGA CTTAATTATT	1860

TGCCTTTTGA AATCACTGTA AATGCCCCCA TCCGGTTCCT CTTCTTCCCA GGTGTGCCAA 1920  
GGAATTAATC TTGGTTTCAC TACAATTAAA ATTCACTCCT TTCCAATCAT GTCATTGAAA 1980  
GTGCCTTTAA CGAAAGAAAT GGTCACTGAA TGGGAATTCT CTTAAGAAAC CCTGAGATTA 2040  
AAAAAAGACT ATTTGGATAA CTTATAGGAA AGCCTAGAAC CTCCCAGTAG AGTGGGGATT 2100  
TTTTTCTTCT TCCCTTTCTC TTTTGGACAA TAGTTAAATT AGCAGTATTA GTTATGAGTT 2160  
TGGTTGCAGT GTTCTTATCT TGTGGGCTGA TTTCCAAAAA CCACATGCTG CTGAATTTAC 2220  
CAGGGATCCT CATACTCAC AATGCAAACC ACTTACTACC AGGCCTTTTT CTGTGTCCAC 2280  
TGGAGAGCTT GAGCTCACAC TCAAAGATCA GAGGACCTAC AGAGAGGGCT CTTTGGTTTG 2340  
AGGACCATGG CTTACCTTTC CTGCCTTTGA CCCATCACAC CCCATTTCCT CCTCTTTCCC 2400  
TCTCCCCGCT GCCAAAAAAA AAAAAAAG GAAACGTTTA TCATGAATCA ACAGGGTTTC 2460  
AGTCCTTATC AAAGAGAGAT GTGGAAAGAG CTAAAGAAAC CACCCTTTGT TCCCAACTCC 2520  
ACTTTACCCA TATTTTATGC AACACAAACA CTGTCCTTTT GGGTCCCTTT CTTACAGATG 2580  
GACCTCTTGA GAAGAATTAT CGTATTCCAC GTTTTTAGCC CTCAGGTAC CAAGATAAAT 2640  
ATATGTATAT ATAACCTTTA TTATTGCTAT ATCTTTGTGG ATAATACATT CAGGTGGTGC 2700  
TGGGTGATTT ATTATAATCT GAACCTAGGT ATATCCTTTG GTCTTCACA GTCATGTTGA 2760  
GGTGGGCTCC CTGGTATGGT AAAAAGCCAG GTATAATGTA ACTTCACCCC AGCCTTTGTA 2820  
CTAAGCTCTT GATAGTGGAT ATACTCTTTT AAGTTTAGCC CCAATATAGG GTAATGAAA 2880  
TTTCTGCCC TCTGGGTTCC CCATTTTAC TATTAAGAAG ACCAGTGATA ATTTAATAAT 2940  
GCCACCAACT CTGGCTTAGT TAAGTGAGAG TGTGAACTGT GTGGCAAGAG AGCCTCACAC 3000  
CTCACTAGGT GCAGAGAGCC CAGGCCTTAT GTTAAAAATCA TGCACCTGAA AAGCAAACCT 3060  
TAATCTGCAA AGACAGCAGC AAGCATTATA CGGTCATCTT GAATGATCCC TTTGAAATTT 3120  
TTTTTTTGTG TGTGTTTAA AATCAAGCCT GAGGCTGGTG AACAGTAGCT ACACACCCAT 3180  
ATTGTGTGTT CTGTGAATGC TAGCTTCTT GAATTTGGAT ATTGGTTATT TTTTATAGAG 3240  
TGTAACCAA GTTTTATATT CTGCAATGCG AACAGGTACC TATCTGTTTC TAAATAAAAC 3300  
TGTTTACATT C 3311

Sequence No.: 21

Sequence length: 1152

Sequence type: Nucleic acid

95

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Stomach cancer

Clone name: HP00876

Sequence characteristics:

Code representing characteristics: CDS

Existence site: 147.. 623

Characterization method: E

Sequence description

ACTGGAGACA CTGAAGAAGG CAGGGGCCCT TAGAGTCTTG GTTGCCAAAC AGATTTGCAG	60
ATCAAGGAGA ACCCAGGAGT TTCAAAGAAG CGCTAGTAAG GTCTCTGAGA TCCTTGCACT	120
AGCTACATCC TCAGGGTAGG AGGAAG ATG GCT TCC AGA AGC ATG CGG CTG CTC	173
Met Ala Ser Arg Ser Met Arg Leu Leu	
1 5	
CTA TTG CTG AGC TGC CTG GCC AAA ACA GGA GTC CTG GGT GAT ATC ATC	221
Leu Leu Leu Ser Cys Leu Ala Lys Thr Gly Val Leu Gly Asp Ile Ile	
10 15 20 25	
ATG AGA CCC AGC TGT GCT CCT GGA TGG TTT TAC CAC AAG TCC AAT TGC	269
Met Arg Pro Ser Cys Ala Pro Gly Trp Phe Tyr His Lys Ser Asn Cys	
30 35 40	
TAT GGT TAC TTC AGG AAG CTG AGG AAC TGG TCT GAT GCC GAG CTC GAG	317
Tyr Gly Tyr Phe Arg Lys Leu Arg Asn Trp Ser Asp Ala Glu Leu Glu	
45 50 55	
TGT CAG TCT TAC GGA AAC GGA GCC CAC CTG GCA TCT ATC CTG AGT TTA	365
Cys Gln Ser Tyr Gly Asn Gly Ala His Leu Ala Ser Ile Leu Ser Leu	
60 65 70	

96

AAG GAA GCC AGC ACC ATA GCA GAG TAC ATA AGT GGC TAT CAG AGA AGC 413  
 Lys Glu Ala Ser Thr Ile Ala Glu Tyr Ile Ser Gly Tyr Gln Arg Ser  
 75 80 85  
 CAG CCG ATA TGG ATT GGC CTG CAC GAC CCA CAG AAG AGG CAG CAG TGG 461  
 Gln Pro Ile Trp Ile Gly Leu His Asp Pro Gln Lys Arg Gln Gln Trp  
 90 95 100 105  
 CAG TGG ATT GAT GGG GCC ATG TAT CTG TAC AGA TCC TGG TCT GGC AAG 509  
 Gln Trp Ile Asp Gly Ala Met Tyr Leu Tyr Arg Ser Trp Ser Gly Lys  
 110 115 120  
 TCC ATG GGT GGG AAC AAG CAC TGT GCT GAG ATG AGC TCC AAT AAC AAC 557  
 Ser Met Gly Gly Asn Lys His Cys Ala Glu Met Ser Ser Asn Asn Asn  
 125 130 135  
 TTT TTA ACT TGG AGC AGC AAC GAA TGC AAC AAG CGC CAA CAC TTC CTG 605  
 Phe Leu Thr Trp Ser Ser Asn Glu Cys Asn Lys Arg Gln His Phe Leu  
 140 145 150  
 TGC AAG TAC CGA CCA TAGAGCAAGA ATCAAGATTC TGCTAACTCC 650  
 Cys Lys Tyr Arg Pro  
 155  
 TGCACAGCCC CGTCCTCTTC CTTTCTGCTA GCCTGGCTAA ATCTGCTCAT TATTTACAGAG 710  
 GGGAAACCTA GCAAACCTAAG AGTGATAAGG GCCCTACTAC ACTGGCTTTT TTAGGCTTAG 770  
 AGACAGAAAC TTTAGCATTG GCCCAGTAGT GGCTTCTAGC TCTAAATGTT TGCCCCGCCA 830  
 TCCCTTTCCA CAGTATCCTT CTTCCCTCCT CCCCTGTCTC TGGCTGTCTC GAGCAGTCTA 890  
 GAAGAGTGCA TCTCCAGCCT ATGAAACAGC TGGGTCTTTG GCCATAAGAA GTAAAGATTT 950  
 GAAGACAGAA GGAAGAACT CAGGAGTAAG CTTCTAGCCC CCTTCAGCTT CTACACCCTT 1010  
 CTGCCCTCTC TCCATTGCCT GCACCCACC CCAGCCACTC AACTCCTGCT TGTTTTCTCT 1070  
 TTGGCCATGG GAAGGTTTAC CAGTAGAATC CTTGCTAGGT TGATGTGGGC CATACATTCC 1130  
 TTTAATAAAC CATTGTGTAC AT 1152

Sequence No.: 22

97

Sequence length: 1749

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Liver

Clone name: HP01134

Sequence characteristics:

Code representing characteristics: CDS

Existence site: 117.. 1247

Characterization method: E

Sequence description

AATCACAGCA GTNCCGACGT CGTGGGTGTT TGGTGTGAGG CTGCGAGCCG CCGCCGCCAC	60
CACTGCCACC ACGGTCGCCT GCCACAGGTG TCTCCAATTG AACTCCAAGG TGCAGA ATG	119
	Met
	1
GTT TGG AAA GTA GCT GTA TTC CTC AGT GTG GCC CTG GGC ATT GGT GCC	167
Val Trp Lys Val Ala Val Phe Leu Ser Val Ala Leu Gly Ile Gly Ala	
5 10 15	
GTT CCT ATA GAT GAT CCT GAA GAT GGA GGC AAG CAC TGG GTG GTG ATC	215
Val Pro Ile Asp Asp Pro Glu Asp Gly Gly Lys His Trp Val Val Ile	
20 25 30	
GTG GCA GGT TCA AAT GGC TGG TAT AAT TAT AGG CAC CAG GCA GAC GCG	263
Val Ala Gly Ser Asn Gly Trp Tyr Asn Tyr Arg His Gln Ala Asp Ala	
35 40 45	
TGC CAT GCC TAC CAG ATC ATT CAC CGC AAT GGG ATT CCT GAC GAA CAG	311
Cys His Ala Tyr Gln Ile Ile His Arg Asn Gly Ile Pro Asp Glu Gln	



98

50	55	60	65	
ATC GTT GTG ATG ATG TAC GAT GAC ATT GCT TAC TCT GAA GAC AAT CCC				359
Ile Val Val Met Met Tyr Asp Asp Ile Ala Tyr Ser Glu Asp Asn Pro				
	70	75	80	
ACT CCA GGA ATT GTG ATC AAC AGG CCC AAT GGC ACA GAT GTC TAT CAG				407
Thr Pro Gly Ile Val Ile Asn Arg Pro Asn Gly Thr Asp Val Tyr Gln				
	85	90	95	
GGA GTC CCG AAG GAC TAC ACT GGA GAG GAT GTT ACC CCA CAA AAT TTC				455
Gly Val Pro Lys Asp Tyr Thr Gly Glu Asp Val Thr Pro Gln Asn Phe				
	100	105	110	
CTT GCT GTG TTG AGA GGC GAT GCA GAA GCA GTG AAG GGC ATA GGA TCC				503
Leu Ala Val Leu Arg Gly Asp Ala Glu Ala Val Lys Gly Ile Gly Ser				
	115	120	125	
GGC AAA GTC CTG AAG AGT GGC CCC CAG GAT CAC GTG TTC ATT TAC TTC				551
Gly Lys Val Leu Lys Ser Gly Pro Gln Asp His Val Phe Ile Tyr Phe				
	130	135	140	145
ACT GAC CAT GGA TCT ACT GGA ATA CTG GTT TTT CCC AAT GAA GAT CTT				599
Thr Asp His Gly Ser Thr Gly Ile Leu Val Phe Pro Asn Glu Asp Leu				
	150	155	160	
CAT GTA AAG GAC CTG AAT GAG ACC ATC CAT TAC ATG TAC AAA CAC AAA				647
His Val Lys Asp Leu Asn Glu Thr Ile His Tyr Met Tyr Lys His Lys				
	165	170	175	
ATG TAC CGA AAG ATG GTG TTC TAC ATT GAA GCC TGT GAG TCT GGG TCC				695
Met Tyr Arg Lys Met Val Phe Tyr Ile Glu Ala Cys Glu Ser Gly Ser				
	180	185	190	
ATG ATG AAC CAC CTG CCG GAT AAC ATC AAT GTT TAT GCA ACT ACT GCT				743
Met Met Asn His Leu Pro Asp Asn Ile Asn Val Tyr Ala Thr Thr Ala				
	195	200	205	
GCC AAC CCC AGA GAG TCG TCC TAC GCC TGT TAC TAT GAT GAG AAG AGG				791

Ala	Asn	Pro	Arg	Glu	Ser	Ser	Tyr	Ala	Cys	Tyr	Tyr	Asp	Glu	Lys	Arg	
210					215				220					225		
TCC	ACG	TAC	CTG	GGG	GAC	TGG	TAC	AGC	GTC	AAC	TGG	ATG	GAA	GAC	TCG	839
Ser	Thr	Tyr	Leu	Gly	Asp	Trp	Tyr	Ser	Val	Asn	Trp	Met	Glu	Asp	Ser	
				230					235					240		
GAC	GTG	GAA	GAT	CTG	ACT	AAA	GAG	ACC	CTG	CAC	AAG	CAG	TAC	CAC	CTG	887
Asp	Val	Glu	Asp	Leu	Thr	Lys	Glu	Thr	Leu	His	Lys	Gln	Tyr	His	Leu	
				245					250					255		
GTA	AAA	TCG	CAC	ACC	AAC	ACC	AGC	CAC	GTC	ATG	CAG	TAT	GGA	AAC	AAA	935
Val	Lys	Ser	His	Thr	Asn	Thr	Ser	His	Val	Met	Gln	Tyr	Gly	Asn	Lys	
				260					265					270		
ACA	ATC	TCC	ACC	ATG	AAA	GTG	ATG	CAG	TTT	CAG	GGT	ATG	AAA	CGC	AAA	983
Thr	Ile	Ser	Thr	Met	Lys	Val	Met	Gln	Phe	Gln	Gly	Met	Lys	Arg	Lys	
				275					280					285		
GCC	AGT	TCT	CCC	GTC	CCC	CTA	CCT	CCA	GTC	ACA	CAC	CTT	GAC	CTC	ACC	1031
Ala	Ser	Ser	Pro	Val	Pro	Leu	Pro	Pro	Val	Thr	His	Leu	Asp	Leu	Thr	
290					295				300					305		
CCC	AGC	CCT	GAT	GTG	CCT	CTC	ACC	ATC	ATG	AAA	AGG	AAA	CTG	ATG	AAC	1079
Pro	Ser	Pro	Asp	Val	Pro	Leu	Thr	Ile	Met	Lys	Arg	Lys	Leu	Met	Asn	
				310					315					320		
ACC	AAT	GAT	CTG	GAG	GAG	TCC	AGG	CAG	CTC	ACG	GAG	GAG	ATC	CAG	CGG	1127
Thr	Asn	Asp	Leu	Glu	Glu	Ser	Arg	Gln	Leu	Thr	Glu	Glu	Ile	Gln	Arg	
				325					330					335		
CAT	CTG	GAT	TAC	GAG	TAT	GCG	TTG	AGA	CAT	TTG	TAC	GTG	CTG	GTC	AAC	1175
His	Leu	Asp	Tyr	Glu	Tyr	Ala	Leu	Arg	His	Leu	Tyr	Val	Leu	Val	Asn	
				340					345					350		
CTT	TGT	GAG	AAG	CCG	TAT	CCG	CTT	CAC	AGG	ATA	AAA	TTG	TCC	ATG	GAC	1223
Leu	Cys	Glu	Lys	Pro	Tyr	Pro	Leu	His	Arg	Ile	Lys	Leu	Ser	Met	Asp	
355					360				365							

100

CAC GTG TGC CTT GGT CAC TAC TGAAGAGCTG CCTCCTGGAA GCTTTT 1270

His Val Cys Leu Gly His Tyr

370

375

CCAAGTGTGA GCGCCCCACC GACTGTGTGC TGATCAGAGA CTGGAGAGGT GGAGTGAGAA 1330

GTCTCCGCTG CTCGGGGCCCT CCTGGGGAGC CCCCCTCCA GGGCTCGCTC CAGGACCTTC 1390

TTCACAAGAT GACTTGCTCG CTGTTACCTG CTTCCCAGT CTTTCTGAA AACTACAAA 1450

TTAGGGTGGG AAAAGCTCTG TATTGAGAAG GGTCAATTTT GCTTCTAGG AGGTTTGTTC 1510

TTTTGCCTGT TAGTTTTGAG GAGCAGGAAG CTCATGGGGG CTTCTGTAGC CCCTCTCAAA 1570

AGGAGTCTTT ATTCTGAGAA TTTGAAGCTG AAACCTCTTT AAATCTTCAG AATGATTTTA 1630

TTGAAGAGGG CCGCAAGCCC CAAATGAAA ACTGTTTTTA GAAAATATGA TGATTTTGA 1690

TTGCTTTTGT ATTTAATTCT GCAGGTGTTT AAGTCTTAAA AAATAAAGAT TTATAACAG 1749

Sequence No.: 23

Sequence length: 988

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Epidermoid carcinoma

Cell line: KB

Clone name: HP10029

Sequence characteristics:

Code representing characteristics: CDS

Existence site: 9.. 530

Characterization method: E

Sequence description

AGTCCAAC ATG GCG GCG CCC AGC GGA GGG TGG AAC GGC GTC CGC GCG AGC 50

101

Met Ala Ala Pro Ser Gly Gly Trp Asn Gly Val Arg Ala Ser

1	5	10	
TTG TGG GCC GCG CTG CTC CTA GGG GCC GTG GCG CTG AGG CCG GCG GAG	98		
Leu Trp Ala Ala Leu Leu Leu Gly Ala Val Ala Leu Arg Pro Ala Glu			
15	20	25	30
GCG GTG TCC GAG CCC ACG ACC GTG GCG TTT GAC GTG CCG CCC GGC GGC	146		
Ala Val Ser Glu Pro Thr Thr Val Ala Phe Asp Val Arg Pro Gly Gly			
35	40	45	
GTC GTG CAT TCC TTC TCC CAT AAC GTG GGC CCG GGG GAC AAA TAT ACG	194		
Val Val His Ser Phe Ser His Asn Val Gly Pro Gly Asp Lys Tyr Thr			
50	55	60	
TGT ATG TTC ACT TAC GCC TCT CAA GGA GGG ACC AAT GAG CAA TGG CAG	242		
Cys Met Phe Thr Tyr Ala Ser Gln Gly Gly Thr Asn Glu Gln Trp Gln			
65	70	75	
ATG AGT CTG GGG ACC AGC GAA GAC CAC CAG CAC TTC ACC TGC ACC ATC	290		
Met Ser Leu Gly Thr Ser Glu Asp His Gln His Phe Thr Cys Thr Ile			
80	85	90	
TGG AGG CCC CAG GGG AAG TCC TAT CTG TAC TTC ACA CAG TTC AAG GCA	338		
Trp Arg Pro Gln Gly Lys Ser Tyr Leu Tyr Phe Thr Gln Phe Lys Ala			
95	100	105	110
GAG GTG CCG GGC GCT GAG ATT GAG TAC GCC ATG GCC TAC TCT AAA GCC	386		
Glu Val Arg Gly Ala Glu Ile Glu Tyr Ala Met Ala Tyr Ser Lys Ala			
115	120	125	
GCA TTT GAA AGG GAA AGT GAT GTC CCT CTG AAA ACT GAG GAA TTT GAA	434		
Ala Phe Glu Arg Glu Ser Asp Val Pro Leu Lys Thr Glu Glu Phe Glu			
130	135	140	
GTG ACC AAA ACA GCA GTG GCT CAC AGG CCC GGG GCA TTC AAA GCT GAG	482		
Val Thr Lys Thr Ala Val Ala His Arg Pro Gly Ala Phe Lys Ala Glu			
145	150	155	

102

CTG TCC AAG CTG GTG ATT GTG GCC AAG GCA TCG CGC ACT GAG CTG	527
Leu Ser Lys Leu Val Ile Val Ala Lys Ala Ser Arg Thr Glu Leu	
160 165 170	
TGA CCAGCAGCCC TGTTCGGGT GGCACCTTCT CATCTCCGGT GAAGCTGAAG	580
GGGCCTGTGG CCCTGAAAGG GCCAGCACAT CACTGGTTTT CTAGGAGGGA CTCTTAAGTT	640
TTCTACCTGG GCTGACGTTG CCTTGTCCGG AGGGGCTTGC AGGGTGGCTG AAGCCCTGGG	700
GCAGAGAACA GAGGGTCCAG GGCCCTCCTG GCTCCCAACA GCTTCTCAGT TCCCACTTCC	760
TGCTGAGCTC TTCTGGACTC AGGATCGCAG ATCCGGGGCA CAAAGAGGGT GGGGAACATG	820
GGGGCTATGC TGGGGAAAGC AGCCATGCTC CCCCCGACCT CCAGCCGAGC ATCCTTCATG	880
AGCCTGCAGA ACTGCTTTCC TATGTTTACC CAGGGGACCT CCTTTCAGAT GAACTGGGAA	940
GAGATGAAAT GTTTTTTCAT ATTTAAATAA ATAAGAACAT TAAAAAGC	988

Sequence No.: 24

Sequence length: 390

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Epidermoid carcinoma

Cell line: KB

Clone name: HP10189

Sequence characteristics:

Code representing characteristics: CDS

Existence site: 102.. 323

Characterization method: E

Sequence description

AATCAGCTTC AGCAATGGAG CGTGCAAAAC ACCAGTGAGC TTCTGTCTTG CTGGAGGGTC	60
-------------------------------------------------------------------	----

103

GGCTTTGGGC GGAAGTGGCT TTGTTGACCG GGAGAAACGA G ATG GGG GTG AAG CTG 116

Met Gly Val Lys Leu

1 5

GAG ATA TTT CGG ATG ATA ATC TAC CTC ACT TTC CCT GTG GCT ATG TTC 164

Glu Ile Phe Arg Met Ile Ile Tyr Leu Thr Phe Pro Val Ala Met Phe

10 15 20

TGG GTT TCC AAT CAG GCC GAG TGG TTT GAG GAC GAT GTC ATA CAG CGC 212

Trp Val Ser Asn Gln Ala Glu Trp Phe Glu Asp Asp Val Ile Gln Arg

25 30 35

AAG AGG GAG CTG TGG CCA CCT GAG AAG CTT CAA GAG ATA GAG GAA TTC 260

Lys Arg Glu Leu Trp Pro Pro Glu Lys Leu Gln Glu Ile Glu Glu Phe

40 45 50

AAA GAG AGG TTA CGG AAG CGG CGG GAG GAG AAG CTC CTT CGC GAC GCC 308

Lys Glu Arg Leu Arg Lys Arg Arg Glu Glu Lys Leu Leu Arg Asp Ala

55 60 65

CAG CAG AAC TCC TGAGGCCTCC AAGTGGGAGT CCTAGCCCCT 350

Gln Gln Asn Ser

70

CCCCTGATGA AATATACATA TACTCAGTTC CTTGTTATTC 390

Sequence No.: 25

Sequence length: 4667

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Lymphoma

104

Cell line: U937

Clone name: HP10269

## Sequence characteristics:

Code representing characteristics: CDS

Existence site: 754.. 4272

Characterization method: E

## Sequence description

CATTAGTTA CTCTGCTCAT TTCTCTTAAG CTTTCCTTGG ATGAGTTGAG CTTTGAATCC 60  
 TTCCTGATGA ACCTTGCCTT TTAAGGATCC TCCAAATGCC CCAAGAAGCT GGGATTTTTC 120  
 ATTTTTTTTT TCACTGGGGA GGGGAATGGT GCTTTCCAGG GTCCTGGATG TTTGAGTCTT 180  
 CTCACCTTCC AGCCCGGTGA TATGTCTGGA GCTTTAACTC TCTATATAAG CCCTAATCTT 240  
 TGTGTTCTCT GCCTGATCTT CTGTCTGGGG TGGTCCAGGT CACAAGAAGA AGCTGACCCC 300  
 TGCTGGCTTT GGGAAAATGC TGAGTTCATT GCCTGGCACA AATGCAAGGG CCCTTCCCGA 360  
 CCCTGTGAAT TCTGGTCTCT GATGATCACT TACATGTGCC TTGTGCTTTC TGTTTGAGGG 420  
 GCCCCTTGCA GCCCCACAG GCAGGTGGGC ATTGTGGAGC TCACTACAAG AACTCTGGGA 480  
 CCGACCGACC AACCCACTTG CCCAGTCCCG TCCTGGGAGG TGGGGGTGCA GTGACGACAG 540  
 ATGGGTGTGA CGGCTGCCAG ATTCTGAGA CCCGCCCTGC GGTGGGGCTA CACCCAGCCA 600  
 GGGAGTCTCC AGAGGTGAGG CTGTTGTTTA AAAACCTGGA GCCGGGAGGG GAGACCCCCA 660  
 CATTCAAGAG GAGCTTTCAG GCGATCTGGA GAAAGAACGG CAGAACACAC AGCAAGGAAA 720  
 GGTCCCTTCT GGGGATCACC CCATTGGCTG AAG ATG AGA CCA TTC TTC CTC TTG 774

Met Arg Pro Phe Phe Leu Leu

1

5

TGT TTT GCC CTG CCT GGC CTC CTG CAT GCC CAA CAA GCC TGC TCC CGT 822  
 Cys Phe Ala Leu Pro Gly Leu Leu His Ala Gln Gln Ala Cys Ser Arg

10

15

20

GGG GCC TGC TAT CCA CCT GTT GGG GAC CTG CTT GTT GGG AGG ACC CGG 870  
 Gly Ala Cys Tyr Pro Pro Val Gly Asp Leu Leu Val Gly Arg Thr Arg

25

30

35

TTT CTC CGA GCT TCA TCT ACC TGT GGA CTG ACC AAG CCT GAG ACC TAC 918

105

Phe	Leu	Arg	Ala	Ser	Ser	Thr	Cys	Gly	Leu	Thr	Lys	Pro	Glu	Thr	Tyr		
40					45					50					55		
TGC	ACC	CAG	TAT	GGC	GAG	TGG	CAG	ATG	AAA	TGC	TGC	AAG	TGT	GAC	TCC	966	
Cys	Thr	Gln	Tyr	Gly	Glu	Trp	Gln	Met	Lys	Cys	Cys	Lys	Cys	Asp	Ser		
				60					65					70			
AGG	CAG	CCT	CAC	AAC	TAC	TAC	AGT	CAC	CGA	GTA	GAG	AAT	GTG	GCT	TCA	1014	
Arg	Gln	Pro	His	Asn	Tyr	Tyr	Ser	His	Arg	Val	Glu	Asn	Val	Ala	Ser		
				75					80					85			
TCC	TCC	GGC	CCC	ATG	CGC	TGG	TGG	CAG	TCC	CAG	AAT	GAT	GTG	AAC	CCT	1062	
Ser	Ser	Gly	Pro	Met	Arg	Trp	Trp	Gln	Ser	Gln	Asn	Asp	Val	Asn	Pro		
				90					95					100			
GTC	TCT	CTG	CAG	CTG	GAC	CTG	GAC	AGG	AGA	TTC	CAG	CTT	CAA	GAA	GTC	1110	
Val	Ser	Leu	Gln	Leu	Asp	Leu	Asp	Arg	Arg	Phe	Gln	Leu	Gln	Glu	Val		
				105					110					115			
ATG	ATG	GAG	TTC	CAG	GGG	CCC	ATG	CCT	GCC	GGC	ATG	CTG	ATT	GAG	CGC	1158	
Met	Met	Glu	Phe	Gln	Gly	Pro	Met	Pro	Ala	Gly	Met	Leu	Ile	Glu	Arg		
120					125					130				135			
TCC	TCA	GAC	TTC	GGT	AAG	ACC	TGG	CGA	GTG	TAC	CAG	TAC	CTG	GCT	GCC	1206	
Ser	Ser	Asp	Phe	Gly	Lys	Thr	Trp	Arg	Val	Tyr	Gln	Tyr	Leu	Ala	Ala		
				140					145					150			
GAC	TGC	ACC	TCC	ACC	TTC	CCT	CGG	GTC	CGC	CAG	GGT	CGG	CCT	CAG	AGC	1254	
Asp	Cys	Thr	Ser	Thr	Phe	Pro	Arg	Val	Arg	Gln	Gly	Arg	Pro	Gln	Ser		
				155					160					165			
TGG	CAG	GAT	GTT	CGG	TGC	CAG	TCC	CTG	CCT	CAG	AGG	CCT	AAT	GCA	CGC	1302	
Trp	Gln	Asp	Val	Arg	Cys	Gln	Ser	Leu	Pro	Gln	Arg	Pro	Asn	Ala	Arg		
				170					175					180			
CTA	AAT	GGG	GGG	AAG	GTC	CAA	CTT	AAC	CTT	ATG	GAT	TTA	GTG	TCT	GGG	1350	
Leu	Asn	Gly	Gly	Lys	Val	Gln	Leu	Asn	Leu	Met	Asp	Leu	Val	Ser	Gly		
				185					190					195			



106

ATT	CCA	GCA	ACT	CAA	AGT	CAA	AAA	ATT	CAA	GAG	GTG	GGG	GAG	ATC	ACA	1398
Ile	Pro	Ala	Thr	Gln	Ser	Gln	Lys	Ile	Gln	Glu	Val	Gly	Glu	Ile	Thr	
200					205					210					215	
AAC	TTG	AGA	GTC	AAT	TTC	ACC	AGG	CTG	GCC	CCT	GTG	CCC	CAA	AGG	GGC	1446
Asn	Leu	Arg	Val	Asn	Phe	Thr	Arg	Leu	Ala	Pro	Val	Pro	Gln	Arg	Gly	
				220					225					230		
TAC	CAC	CCT	CCC	AGC	GCC	TAC	TAT	GCT	GTG	TCC	CAG	CTC	CGT	CTG	CAG	1494
Tyr	His	Pro	Pro	Ser	Ala	Tyr	Tyr	Ala	Val	Ser	Gln	Leu	Arg	Leu	Gln	
				235				240						245		
GGG	AGC	TGC	TTC	TGT	CAC	GGC	CAT	GCT	GAT	CGC	TGC	GCA	CCC	AAG	CCT	1542
Gly	Ser	Cys	Phe	Cys	His	Gly	His	Ala	Asp	Arg	Cys	Ala	Pro	Lys	Pro	
				250				255					260			
GGG	GCC	TCT	GCA	GGC	CCC	TCC	ACC	GCT	GTG	CAG	GTG	CAC	GAT	GTG	TGT	1590
Gly	Ala	Ser	Ala	Gly	Pro	Ser	Thr	Ala	Val	Gln	Val	His	Asp	Val	Cys	
				265				270					275			
GTC	TGC	CAG	CAC	AAC	ACT	GCC	GGC	CCA	AAT	TGT	GAG	CGC	TGT	GCA	CCC	1638
Val	Cys	Gln	His	Asn	Thr	Ala	Gly	Pro	Asn	Cys	Glu	Arg	Cys	Ala	Pro	
280					285					290					295	
TTC	TAC	AAC	AAC	CGG	CCC	TGG	AGA	CCG	GCG	GAG	GGC	CAG	GAC	GCC	CAT	1686
Phe	Tyr	Asn	Asn	Arg	Pro	Trp	Arg	Pro	Ala	Glu	Gly	Gln	Asp	Ala	His	
					300					305				310		
GAA	TGC	CAA	AGG	TGC	GAC	TGC	AAT	GGG	CAC	TCA	GAG	ACA	TGT	CAC	TTT	1734
Glu	Cys	Gln	Arg	Cys	Asp	Cys	Asn	Gly	His	Ser	Glu	Thr	Cys	His	Phe	
					315					320				325		
GAC	CCC	GCT	GTG	TTT	GCC	GCC	AGC	CAG	GGG	GCA	TAT	GGA	GGT	GTG	TGT	1782
Asp	Pro	Ala	Val	Phe	Ala	Ala	Ser	Gln	Gly	Ala	Tyr	Gly	Gly	Val	Cys	
					330					335				340		
GAC	AAT	TGC	CGG	GAC	CAC	ACC	GAA	GGC	AAG	AAC	TGT	GAG	CGG	TGT	CAG	1830
Asp	Asn	Cys	Arg	Asp	His	Thr	Glu	Gly	Lys	Asn	Cys	Glu	Arg	Cys	Gln	

107

345	350	355	
CTG CAC TAT TTC CGG AAC CGG CGC CCG GGA GCT TCC ATT CAG GAG ACC	1878		
Leu His Tyr Phe Arg Asn Arg Arg Pro Gly Ala Ser Ile Gln Glu Thr			
360	365	370	375
TGC ATC TCC TGC GAG TGT GAT CCG GAT GGG GCA GTG CCA GGG GCT CCC	1926		
Cys Ile Ser Cys Glu Cys Asp Pro Asp Gly Ala Val Pro Gly Ala Pro			
380	385	390	
TGT GAC CCA GTG ACC GGG CAG TGT GTG TGC AAG GAG CAT GTG CAG GGA	1974		
Cys Asp Pro Val Thr Gly Gln Cys Val Cys Lys Glu His Val Gln Gly			
395	400	405	
GAG CGC TGT GAC CTA TGC AAG CCG GGC TTC ACT GGA CTC ACC TAC GCC	2022		
Glu Arg Cys Asp Leu Cys Lys Pro Gly Phe Thr Gly Leu Thr Tyr Ala			
410	415	420	
AAC CCG CAG GGC TGC CAC CGC TGT GAC TGC AAC ATC CTG GGG TCC CGG	2070		
Asn Pro Gln Gly Cys His Arg Cys Asp Cys Asn Ile Leu Gly Ser Arg			
425	430	435	
AGG GAC ATG CCG TGT GAC GAG GAG AGT GGG CGC TGC CTT TGT CTG CCC	2118		
Arg Asp Met Pro Cys Asp Glu Glu Ser Gly Arg Cys Leu Cys Leu Pro			
440	445	450	455
AAC GTG GTG GGT CCC AAA TGT GAC CAG TGT GCT CCC TAC CAC TGG AAG	2166		
Asn Val Val Gly Pro Lys Cys Asp Gln Cys Ala Pro Tyr His Trp Lys			
460	465	470	
CTG GCC AGT GGC CAG GGC TGT GAA CCG TGT GCC TGC GAC CCG CAC AAC	2214		
Leu Ala Ser Gly Gln Gly Cys Glu Pro Cys Ala Cys Asp Pro His Asn			
475	480	485	
TCC CTC AGC CCA CAG TGC AAC CAG TTC ACA GGG CAG TGC CCC TGT CGG	2262		
Ser Leu Ser Pro Gln Cys Asn Gln Phe Thr Gly Gln Cys Pro Cys Arg			
490	495	500	
GAA GGC TTT GGT GGC CTG ATG TGC AGC GCT GCA GCC ATC CGC CAG TGT	2310		

108

Glu Gly Phe Gly Gly Leu Met Cys Ser Ala Ala Ala Ile Arg Gln Cys  
 505 510 515  
 CCA GAC CGG ACC TAT GGA GAC GTG GCC ACA GGA TGC CGA GCC TGT GAC 2358  
 Pro Asp Arg Thr Tyr Gly Asp Val Ala Thr Gly Cys Arg Ala Cys Asp  
 520 525 530 535  
 TGT GAT TTC CGG GGA ACA GAG GGC CCG GGC TGC GAC AAG GCA TCA GGC 2406  
 Cys Asp Phe Arg Gly Thr Glu Gly Pro Gly Cys Asp Lys Ala Ser Gly  
 540 545 550  
 CGC TGC CTC TGC CGC CCT GGC TTG ACC GGG CCC CGC TGT GAC CAG TGC 2454  
 Arg Cys Leu Cys Arg Pro Gly Leu Thr Gly Pro Arg Cys Asp Gln Cys  
 555 560 565  
 CAG CGA GGC TAC TGC AAT CGC TAC CCG GTG TGC GTG GCC TGC CAC CCT 2502  
 Gln Arg Gly Tyr Cys Asn Arg Tyr Pro Val Cys Val Ala Cys His Pro  
 570 575 580  
 TGC TTC CAG ACC TAT GAT GCG GAC CTC CGG GAG CAG GCC CTG CGC TTT 2550  
 Cys Phe Gln Thr Tyr Asp Ala Asp Leu Arg Glu Gln Ala Leu Arg Phe  
 585 590 595  
 GGT AGA CTC CGC AAT GCC ACC GCC AGC CTG TGG TCA GGG CCT GGG CTG 2598  
 Gly Arg Leu Arg Asn Ala Thr Ala Ser Leu Trp Ser Gly Pro Gly Leu  
 600 605 610 615  
 GAG GAC CGT GGC CTG GCC TCC CGG ATC CTA GAT GCA AAG AGT AAG ATT 2646  
 Glu Asp Arg Gly Leu Ala Ser Arg Ile Leu Asp Ala Lys Ser Lys Ile  
 620 625 630  
 GAG CAG ATC CGA GCA GTT CTC AGC AGC CCC GCA GTC ACA GAG CAG GAG 2694  
 Glu Gln Ile Arg Ala Val Leu Ser Ser Pro Ala Val Thr Glu Gln Glu  
 635 640 645  
 GTG GCT CAG GTG GCC AGT GCC ATC CTC TCC CTC AGG CGA ACT CTC CAG 2742  
 Val Ala Gln Val Ala Ser Ala Ile Leu Ser Leu Arg Arg Thr Leu Gln  
 650 655 660

109

GGC CTG CAG CTG GAT CTG CCC CTG GAG GAG GAG ACG TTG TCC CTT CCG	2790
Gly Leu Gln Leu Asp Leu Pro Leu Glu Glu Glu Thr Leu Ser Leu Pro	
665 670 675	
AGA GAC CTG GAG AGT CTT GAC AGA AGC TTC AAT GGT CTC CTT ACT ATG	2838
Arg Asp Leu Glu Ser Leu Asp Arg Ser Phe Asn Gly Leu Leu Thr Met	
680 685 690 695	
TAT CAG AGG AAG AGG GAG CAG TTT GAA AAA ATA AGC AGT GCT GAT CCT	2886
Tyr Gln Arg Lys Arg Glu Gln Phe Glu Lys Ile Ser Ser Ala Asp Pro	
700 705 710	
TCA GGA GCC TTC CGG ATG CTG AGC ACA GCC TAC GAG CAG TCA GCC CAG	2934
Ser Gly Ala Phe Arg Met Leu Ser Thr Ala Tyr Glu Gln Ser Ala Gln	
715 720 725	
GCT GCT CAG CAG GTC TCC GAC AGC TCG CGC CTT TTG GAC CAG CTC AGG	2982
Ala Ala Gln Gln Val Ser Asp Ser Ser Arg Leu Leu Asp Gln Leu Arg	
730 735 740	
GAC AGC CGG AGA GAG GCA GAG AGG CTG GTG CGG CAG GCG GGA GGA GGA	3030
Asp Ser Arg Arg Glu Ala Glu Arg Leu Val Arg Gln Ala Gly Gly Gly	
745 750 755	
GGA GGC ACC GGC AGC CCC AAG CTT GTG GCC CTG AGG CTG GAG ATG TCT	3078
Gly Gly Thr Gly Ser Pro Lys Leu Val Ala Leu Arg Leu Glu Met Ser	
760 765 770 775	
TCG TTG CCT GAC CTG ACA CCC ACC TTC AAC AAG CTC TGT GGC AAC TCC	3126
Ser Leu Pro Asp Leu Thr Pro Thr Phe Asn Lys Leu Cys Gly Asn Ser	
780 785 790	
AGG CAG ATG GCT TGC ACC CCA ATA TCA TGC CCT GGT GAG CTA TGT CCC	3174
Arg Gln Met Ala Cys Thr Pro Ile Ser Cys Pro Gly Glu Leu Cys Pro	
795 800 805	
CAA GAC AAT GGC ACA GCC TGT GGC TCC CGC TGC AGG GGT GTC CTT CCC	3222
Gln Asp Asn Gly Thr Ala Cys Gly Ser Arg Cys Arg Gly Val Leu Pro	

110

810	815	820	
AGG GCC GGT GGG GCC TTC TTG ATG GCG GGG CAG GTG GCT GAG CAG CTG			3270
Arg Ala Gly Gly Ala Phe Leu Met Ala Gly Gln Val Ala Glu Gln Leu			
825	830	835	
CGG GGC TTC AAT GCC CAG CTC CAG CGG ACC AGG CAG ATG ATT AGG GCA			3318
Arg Gly Phe Asn Ala Gln Leu Gln Arg Thr Arg Gln Met Ile Arg Ala			
840	845	850	855
GCC GAG GAA TCT GCC TCA CAG ATT CAA TCC AGT GCC CAG CGC TTG GAG			3366
Ala Glu Glu Ser Ala Ser Gln Ile Gln Ser Ser Ala Gln Arg Leu Glu			
860	865	870	
ACC CAG GTG AGC GCC AGC CGC TCC CAG ATG GAG GAA GAT GTC AGA CGC			3414
Thr Gln Val Ser Ala Ser Arg Ser Gln Met Glu Glu Asp Val Arg Arg			
875	880	885	
ACA CGG CTC CTA ATC CAG CAG GTC CGG GAC TTC CTA ACA GAC CCC GAC			3462
Thr Arg Leu Leu Ile Gln Gln Val Arg Asp Phe Leu Thr Asp Pro Asp			
890	895	900	
ACT GAT GCA GCC ACT ATC CAG GAG GTC AGC GAG GCC GTG CTG GCC CTG			3510
Thr Asp Ala Ala Thr Ile Gln Glu Val Ser Glu Ala Val Leu Ala Leu			
905	910	915	
TGG CTG CCC ACA GAC TCA GCT ACT GTT CTG CAG AAG ATG AAT GAG ATC			3558
Trp Leu Pro Thr Asp Ser Ala Thr Val Leu Gln Lys Met Asn Glu Ile			
920	925	930	935
CAG GCC ATT GCA GCC AGG CTC CCC AAC GTG GAC TTG GTG CTG TCC CAG			3606
Gln Ala Ile Ala Ala Arg Leu Pro Asn Val Asp Leu Val Leu Ser Gln			
940	945	950	
ACC AAG CAG GAC ATT GCG CGT GCC CGC CGG TTG CAG GCT GAG GCT GAG			3654
Thr Lys Gln Asp Ile Ala Arg Ala Arg Arg Leu Gln Ala Glu Ala Glu			
955	960	965	
GAA GCC AGG AGC CGA GCC CAT GCA GTG GAG GGC CAG GTG GAA GAT GTG			3702

111

Glu Ala Arg Ser Arg Ala His Ala Val Glu Gly Gln Val Glu Asp Val  
 970 975 980  
 GTT GGG AAC CTG CGG CAG GGG ACA GTG GCA CTG CAG GAA GCT CAG GAC 3750  
 Val Gly Asn Leu Arg Gln Gly Thr Val Ala Leu Gln Glu Ala Gln Asp  
 985 990 995  
 ACC ATG CAA GGC ACC AGC CGC TCC CTT CGG CTT ATC CAG GAC AGG GTT 3798  
 Thr Met Gln Gly Thr Ser Arg Ser Leu Arg Leu Ile Gln Asp Arg Val  
 1000 1005 1010 1015  
 GCT GAG GTT CAG CAG GTA CTG CGG CCA GCA GAA AAG CTG GTG ACA AGC 3846  
 Ala Glu Val Gln Gln Val Leu Arg Pro Ala Glu Lys Leu Val Thr Ser  
 1020 1025 1030  
 ATG ACC AAG CAG CTG GGT GAC TTC TGG ACA CGG ATG GAG GAG CTC CGC 3894  
 Met Thr Lys Gln Leu Gly Asp Phe Trp Thr Arg Met Glu Glu Leu Arg  
 1035 1040 1045  
 CAC CAA GCC CGG CAG CAG GGG GCA GAG GCA GTC CAG GCC CAG CAG CTT 3942  
 His Gln Ala Arg Gln Gln Gly Ala Glu Ala Val Gln Ala Gln Gln Leu  
 1050 1055 1060  
 GCG GAA GGT GCC AGC GAG CAG GCA TTG AGT GCC CAA GAG GGA TTT GAG 3990  
 Ala Glu Gly Ala Ser Glu Gln Ala Leu Ser Ala Gln Glu Gly Phe Glu  
 1065 1070 1075  
 AGA ATA AAA CAA AAG TAT GCT GAG TTG AAG GAC CGG TTG GGT CAG AGT 4038  
 Arg Ile Lys Gln Lys Tyr Ala Glu Leu Lys Asp Arg Leu Gly Gln Ser  
 1080 1085 1090 1095  
 TCC ATG CTG GGT GAG CAG GGT GCC CGG ATC CAG AGT GTG AAG ACA GAG 4086  
 Ser Met Leu Gly Glu Gln Gly Ala Arg Ile Gln Ser Val Lys Thr Glu  
 1100 1105 1110  
 GCA GAG GAG CTG TTT GGG GAG ACC ATG GAG ATG ATG GAC AGG ATG AAA 4134  
 Ala Glu Glu Leu Phe Gly Glu Thr Met Glu Met Met Asp Arg Met Lys  
 1115 1120 1125

112

GAC ATG GAG TTG GAG CTG CTG CGG GGC AGC CAG GCC ATC ATG CTG CGC 4182  
Asp Met Glu Leu Glu Leu Leu Arg Gly Ser Gln Ala Ile Met Leu Arg  
1130 1135 1140  
TCA GCG GAC CTG ACA GGA CTG GAG AAG CGT GTG GAG CAG ATC CGT GAC 4230  
Ser Ala Asp Leu Thr Gly Leu Glu Lys Arg Val Glu Gln Ile Arg Asp  
1145 1150 1155  
CAC ATC AAT GGG CGC GTG CTC TAC TAT GCC ACC TGC AAG T 4270  
His Ile Asn Gly Arg Val Leu Tyr Tyr Ala Thr Cys Lys  
1160 1165 1170  
GATGCTACAG CTTCCAGCCC GTTGCCCCAC TCATCTGCCG CCTTTGCTTT TGGTTGGGGG 4330  
CAGATTGGGT TGGAAATGCTT TCCATCTCCA GGAGACTTTC ATGCAGCCTA AAGTACAGCC 4390  
TGGACCACCC CTGGTGTGTA GCTAGTAAGA TTACCCTGAG CTGCAGCTGA GCCTGAGCCA 4450  
ATGGGACAGT TACACTTGAC AGACAAAGAT GGTGGAGATT GGCATGCCAT TGAAACTAAG 4510  
AGCTCTCAAG TCAAGGAAGC TGGGCTGGGC AGTATCCCCC GCCTTTAGTT CTCCACTGGG 4570  
GAGGAATCCT GGACCAAGCA CAAAACTTA AAAAAAGTGA TGTAAAAATG AAAAGCCAAA 4630  
TAAAAATCTT TGGAAAAGAG CCTGGAGGTT CAACGAG 4667

Sequence No.: 26

Sequence length: 1086

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Stomach cancer

Clone name: HP10298

Sequence characteristics:

Code representing characteristics: CDS

113

Existence site: 138.. 506

Characterization method: E

## Sequence description

TTTAATTTCC CCGAAATCAG ACTGCTGCCT TGGACCGGGA CAGCTCGCGG CCCCCGAGAG	60
CTCTAGCCGT CGAGGAGCTG CCTGGGGACG TTTGCCCTGG GGCCCCAGCC TGGCCCCGGGT	120
CACCCTGGCA TGAGGAG ATG GGC CTG TTG CTC CTG GTC CCA TTG CTC CTG	170
Met Gly Leu Leu Leu Leu Val Pro Leu Leu Leu	
1 5 10	
CTG CCC GGC TCC TAC GGA CTG CCC TTC TAC AAC GGC TTC TAC TAC TCC	218
Leu Pro Gly Ser Tyr Gly Leu Pro Phe Tyr Asn Gly Phe Tyr Tyr Ser	
15 20 25	
AAC AGC GCC AAC GAC CAG AAC CTA GGC AAC GGT CAT GGC AAA GAC CTC	266
Asn Ser Ala Asn Asp Gln Asn Leu Gly Asn Gly His Gly Lys Asp Leu	
30 35 40	
CTT AAT GGA GTG AAG CTG GTG GTG GAG ACA CCC GAG GAG ACC CTG TTC	314
Leu Asn Gly Val Lys Leu Val Val Glu Thr Pro Glu Glu Thr Leu Phe	
45 50 55	
ACC CGC ATC CTA ACT GTG GGC CCC CAG AGC CTG GGG TCC GAA GCT TTG	362
Thr Arg Ile Leu Thr Val Gly Pro Gln Ser Leu Gly Ser Glu Ala Leu	
60 65 70 75	
GCT TCC CCG ACC CGC AGA GCC GCT TGT ACG GTG TTT ACT GCT ACC GCC	410
Ala Ser Pro Thr Arg Arg Ala Ala Cys Thr Val Phe Thr Ala Thr Ala	
80 85 90	
AGC ACT AGG ACC TGG GGC CCT CCC CTG CCG CAT TCC CTC ACT GGC TGT	458
Ser Thr Arg Thr Trp Gly Pro Pro Leu Pro His Ser Leu Thr Gly Cys	
95 100 105	
GTA TTT ATT GAG TGG TTC GTT TTC CCT TGT GGG TTG GAG CCA TTT	503
Val Phe Ile Glu Trp Phe Val Phe Pro Cys Gly Leu Glu Pro Phe	
110 115 120	



114

TAAGTGT TTTTATACTT CTCAATTAA ATTTTCTTAA AACATTTTTT TACTATTTTT 560  
TGTAAGCAA ACAGAACCCA ATGCCTCCCT TTGCTCCTGG ATGCCCCACT CCAGGAATCA 620  
TGCTTGCTCC CCTGGGCCAT TTGCGGTTTT GTGGGCTTCT GGAGGGTTCC CCGCCATCCA 680  
GGCTGGTCTC CCTCCCTTAA GGAGGTGGT GCCCAGAGTG GCGGTGGCC TGTCTAGAAT 740  
GCCGCCGGA GTCCGGGCAT GGTGGGCACA GTTCTCCCTG CCCCTCAGCC TGGGGGAAGA 800  
AGAGGGCCTC GGGGGCCTCC GGAGCTGGGC TTTGGGCCTC TCCTGCCAC CTCTACTTCT 860  
CTGTGAAGCC GCTGACCCCA GTCTGCCAC TGAGGGGCTA GGGCTGGAAG CCAGTTCTAG 920  
GCTTCCAGGC GAAAGCTGAG GGAAGGAAGA AACTCCCCTC CCCGTTCCCC TTCCCCTCTC 980  
GGTTCCAAAG AATCTGTTTT GTTGTCAATT GTTCTCCTG TTTCCCTGTG TGGGGAGGGG 1040  
CCCTCAGGTG TGTGTACTTT GGACAATAAA TGGTGCTATG ACTGCC 1086

Sequence No.: 27

Sequence length: 866

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Stomach cancer

Clone name: HP10368

Sequence characteristics:

Code representing characteristics: CDS

Existence site: 73.. 600

Characterization method: E

Sequence description

ACTCAGAAGC TTGGACCGCA TCCTAGCCGC CGACTCACAC AAGGCAGGTG GGTGAGGAAA 60  
TCCAGAGTTG CC ATG GAG AAA ATT CCA GTG TCA GCA TTC TTG CTC CTT GTG 111

Met Glu Lys Ile Pro Val Ser Ala Phe Leu Leu Leu Val

115

	1	5	10	
GCC CTC TCC TAC ACT CTG GCC AGA GAT ACC ACA GTC AAA CCT GGA GCC				159
Ala Leu Ser Tyr Thr Leu Ala Arg Asp Thr Thr Val Lys Pro Gly Ala				
15	20	25		
AAA AAG GAC ACA AAG GAC TCT CGA CCC AAA CTG CCC CAG ACC CTC TCC				207
Lys Lys Asp Thr Lys Asp Ser Arg Pro Lys Leu Pro Gln Thr Leu Ser				
30	35	40	45	
AGA GGT TGG GGT GAC CAA CTC ATC TGG ACT CAG ACA TAT GAA GAA GCT				255
Arg Gly Trp Gly Asp Gln Leu Ile Trp Thr Gln Thr Tyr Glu Glu Ala				
50	55	60		
CTA TAT AAA TCC AAG ACA AGC AAC AAA CCC TTG ATG ATT ATT CAT CAC				303
Leu Tyr Lys Ser Lys Thr Ser Asn Lys Pro Leu Met Ile Ile His His				
65	70	75		
TTG GAT GAG TGC CCA CAC AGT CAA GCT TTA AAG AAA GTG TTT GCT GAA				351
Leu Asp Glu Cys Pro His Ser Gln Ala Leu Lys Lys Val Phe Ala Glu				
80	85	90		
AAT AAA GAA ATC CAG AAA TTG GCA GAG CAG TTT GTC CTC CTC AAT CTG				399
Asn Lys Glu Ile Gln Lys Leu Ala Glu Gln Phe Val Leu Leu Asn Leu				
95	100	105		
GTT TAT GAA ACA ACT GAC AAA CAC CTT TCT CCT GAT GGC CAG TAT GTC				447
Val Tyr Glu Thr Thr Asp Lys His Leu Ser Pro Asp Gly Gln Tyr Val				
110	115	120	125	
CCC AGG ATT ATG TTT GTT GAC CCA TCT CTG ACA GTT AGA GCC GAT ATC				495
Pro Arg Ile Met Phe Val Asp Pro Ser Leu Thr Val Arg Ala Asp Ile				
130	135	140		
ACT GGA AGA TAT TCA AAC CGT CTC TAT GCT TAC GAA CCT GCA GAT ACA				543
Thr Gly Arg Tyr Ser Asn Arg Leu Tyr Ala Tyr Glu Pro Ala Asp Thr				
145	150	155		
GCT CTG TTG CTT GAC AAC ATG AAG AAA GCT CTC AAG TTG CTG AAG ACT				591

116

Ala Leu Leu Leu Asp Asn Met Lys Lys Ala Leu Lys Leu Leu Lys Thr

160

165

170

GAA TTG TAAAGAAAAA AAATCTCCAA GCCCTTCTGT CTGTCAGGCC TTG

640

Glu Leu

175

AGACTTGAAA CCAGAAGAAG TGTGAGAAGA CTGGCTAGTG TGGAAGCATA GTGAACACAC

700

TGATTAGGTT ATGGTTTAAT GTTACAACAA CTATTTTTTA AGAAAAACAA GTTTTAGAAA

760

TTTGGTTTCA AGTGATCATG TGTGAAAACA ATATTGTATA CTACCATAGT GAGCCATGAT

820

TTTCTAAAAA AAAAAATAAA TGTTTTGGGG GTGTTCTGTT TTCTCC

866

## Claims

1. Proteins containing any of the amino acid sequences represented by Sequence No. 1 to Sequence No. 9.
2. DNAs encoding any of the proteins as described in Claim 1.
3. cDNAs containing any of the base sequences represented by Sequence No. 10 to Sequence No. 18.
4. cDNAs described in Claim 3 which comprise any of the base sequences represented by Sequence No. 19 to Sequence No. 27.

1 / 11

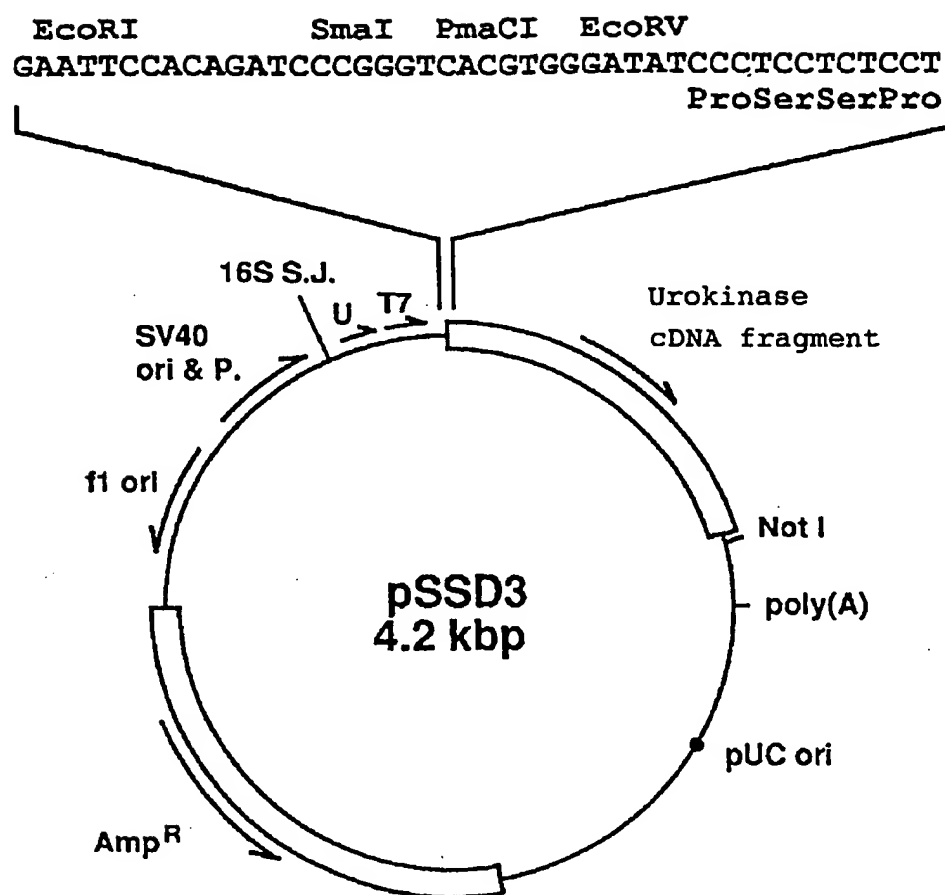
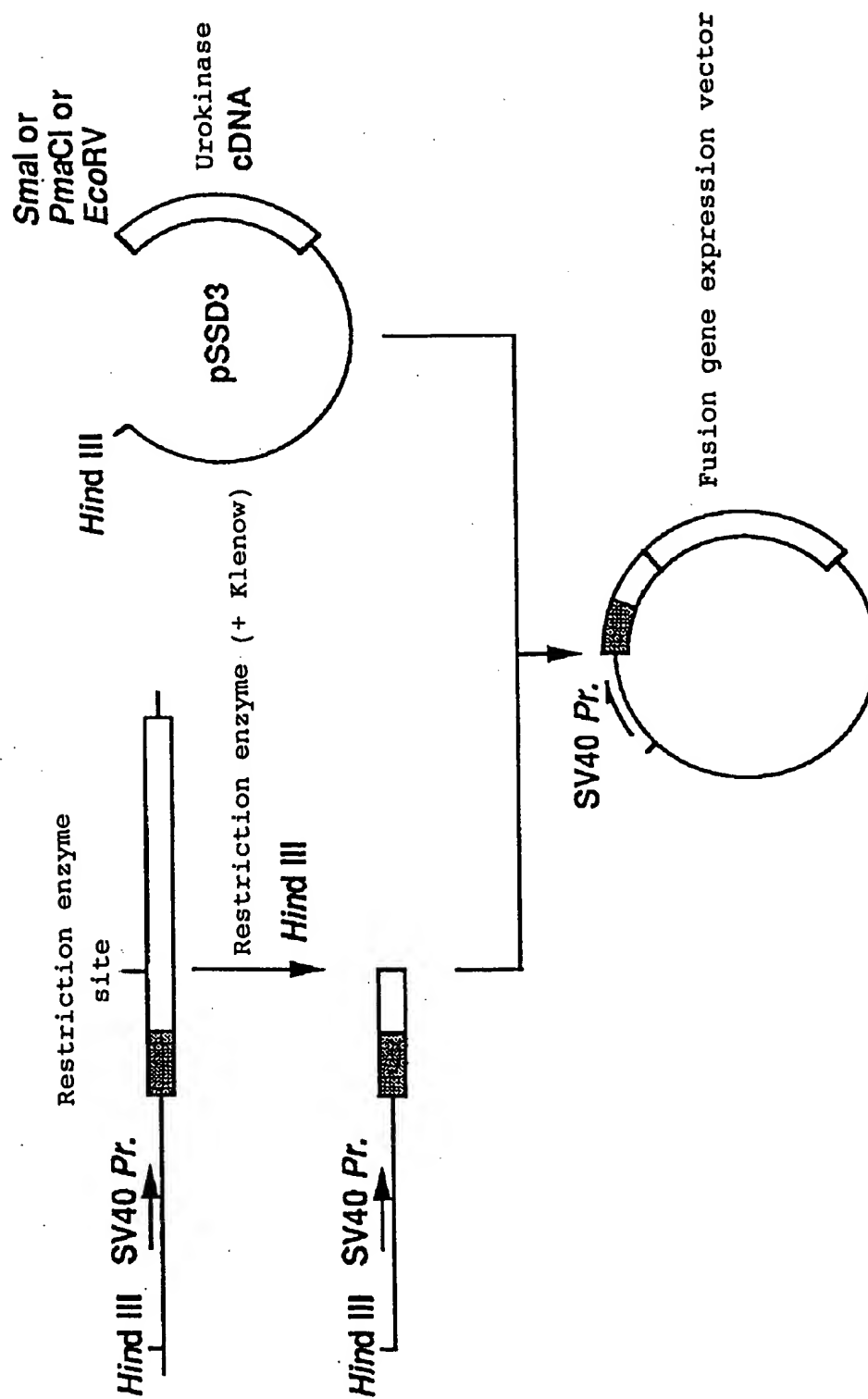


Fig.1

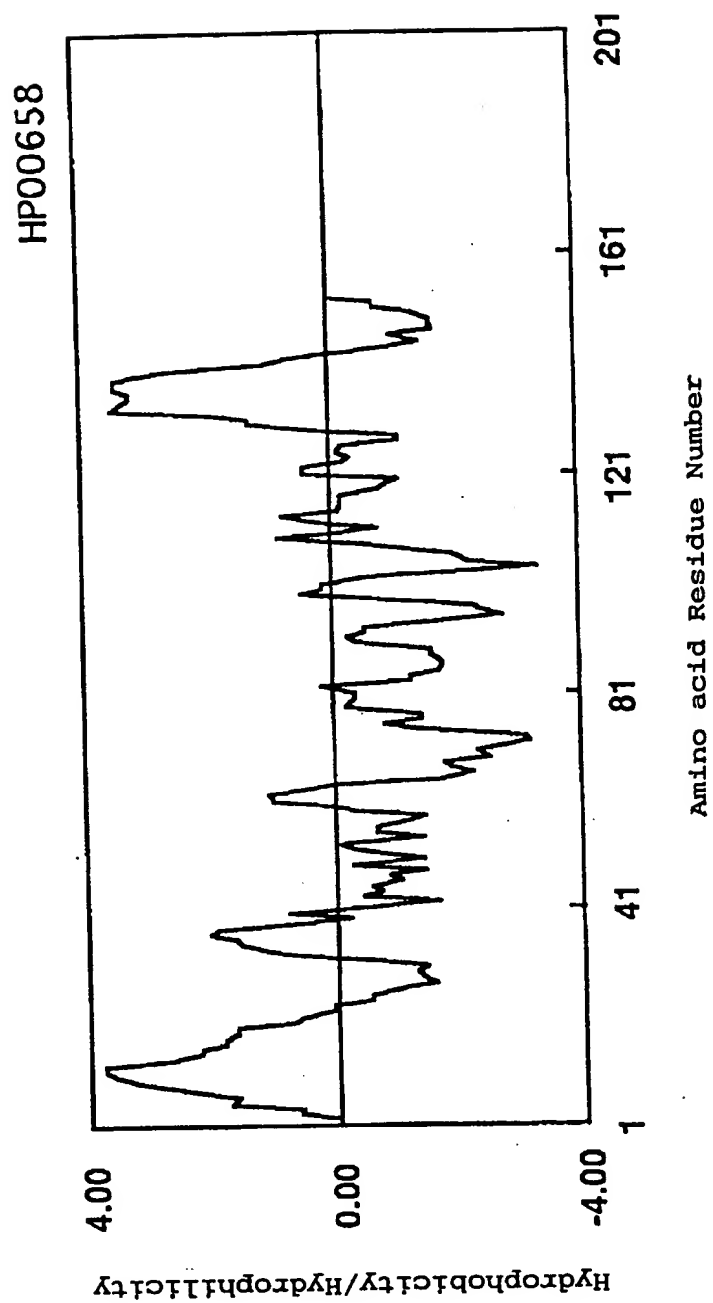
2/11

Fig. 2



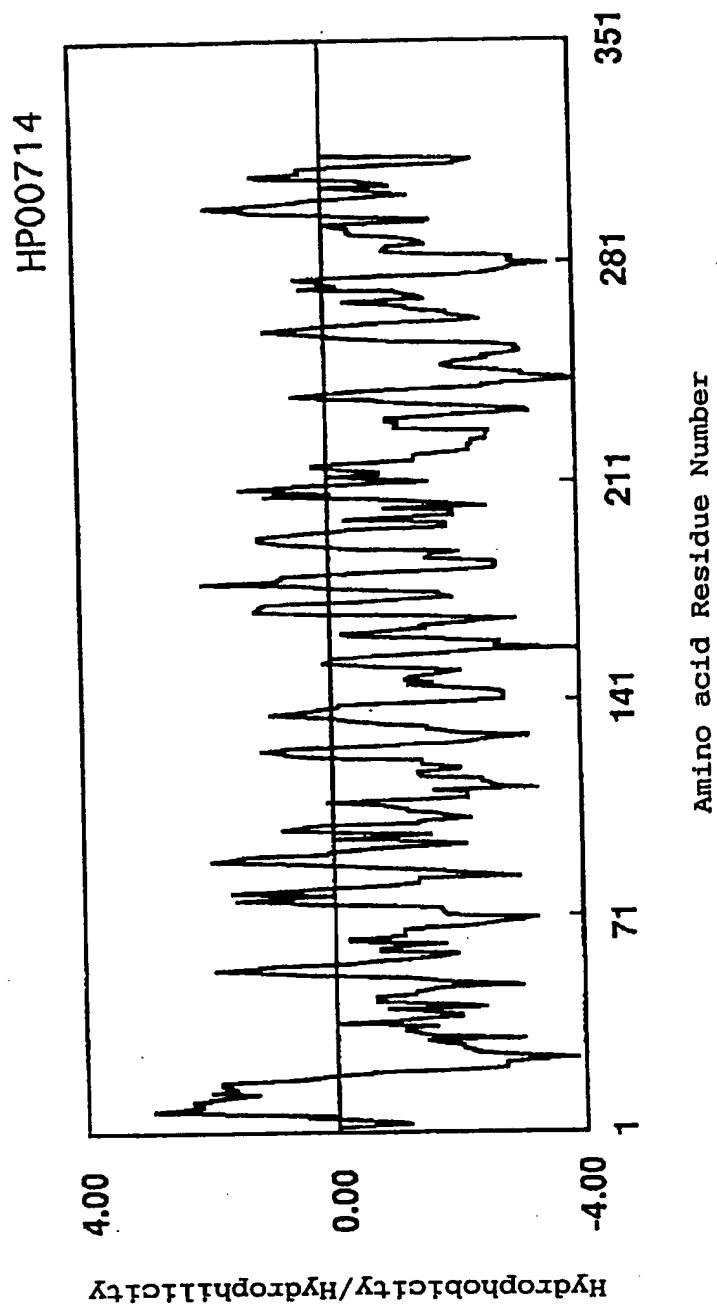
3/11

Fig. 3



4/11

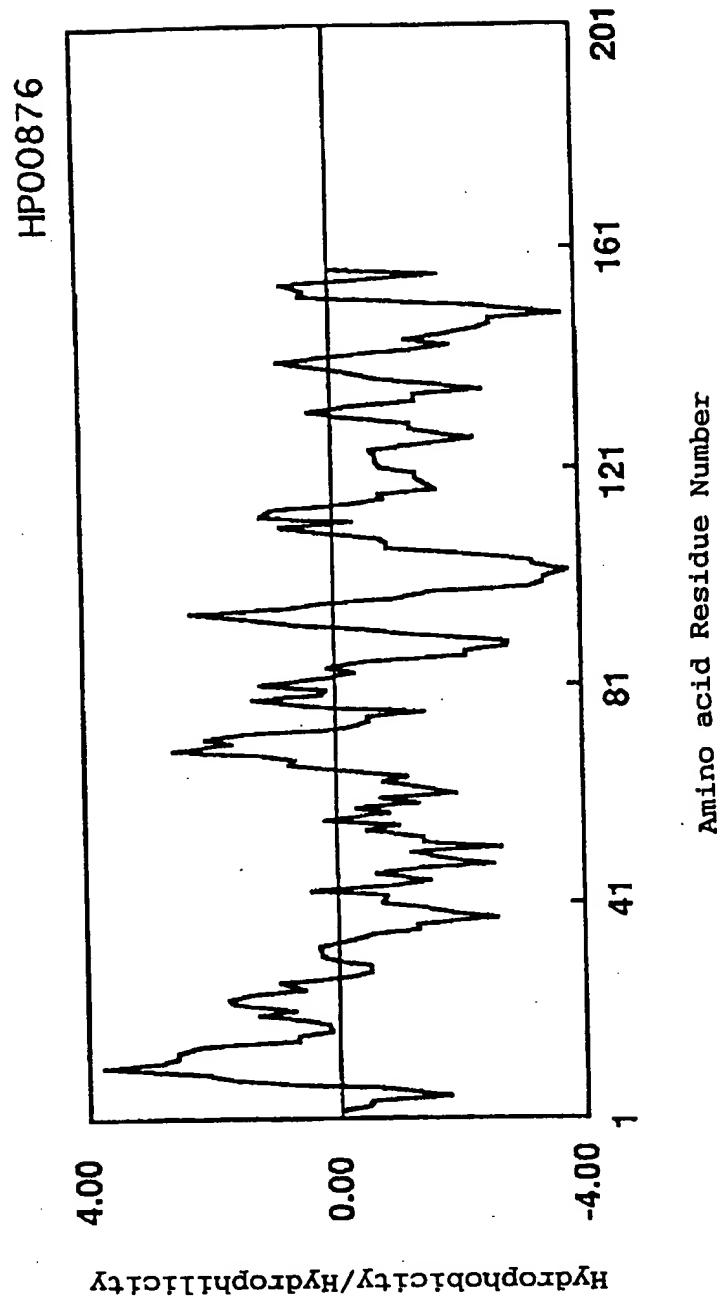
Fig. 4





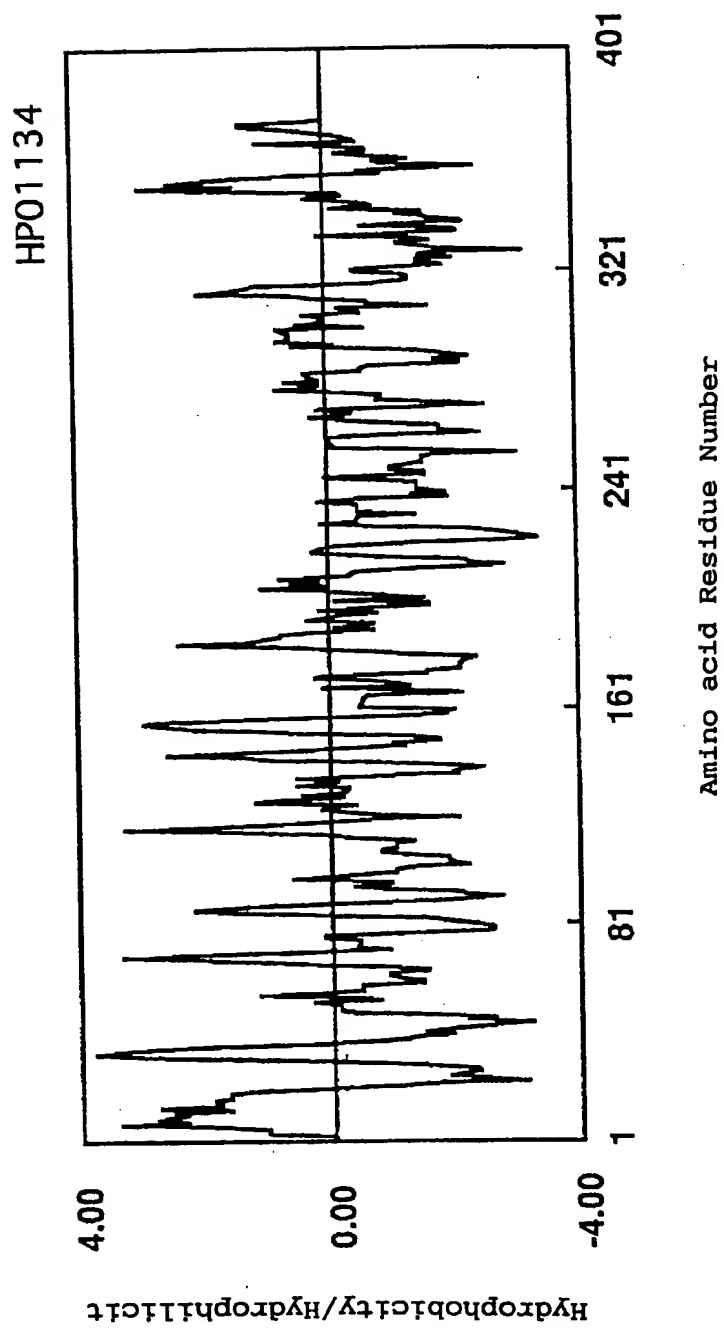
5/11

Fig. 5



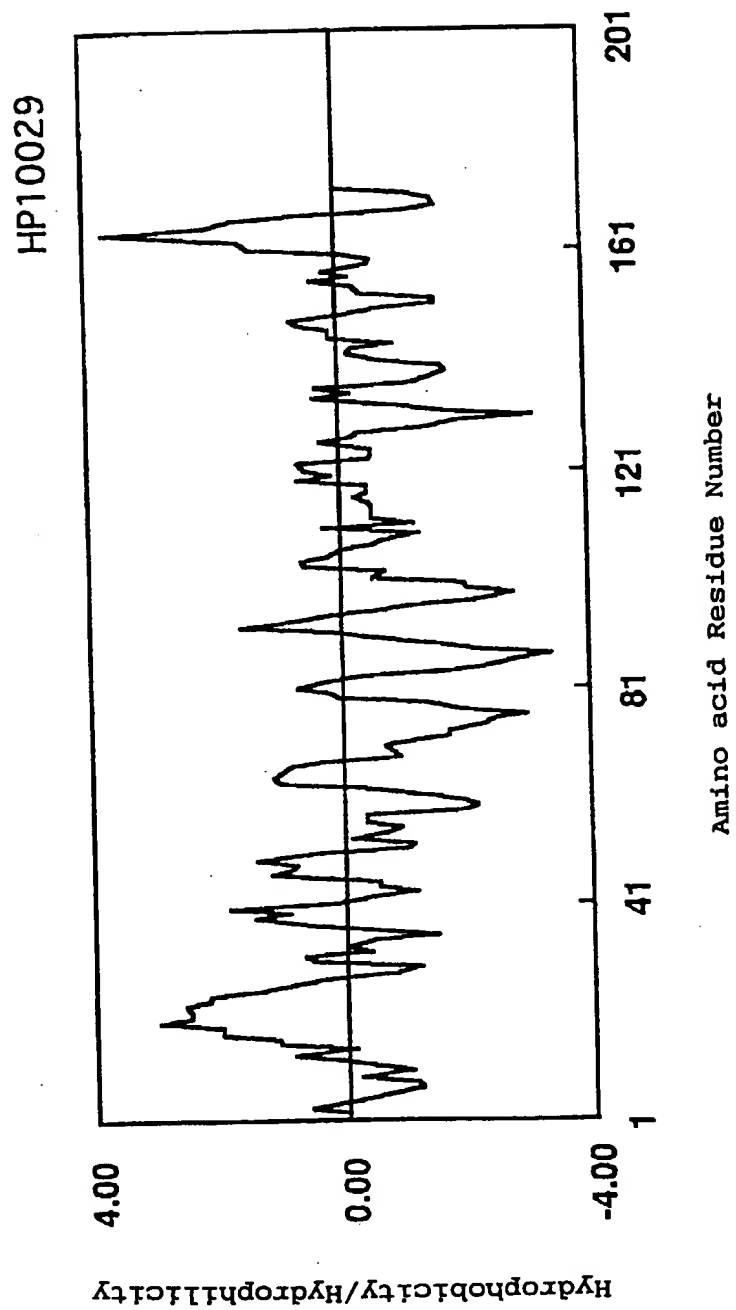
6/11

Fig. 6



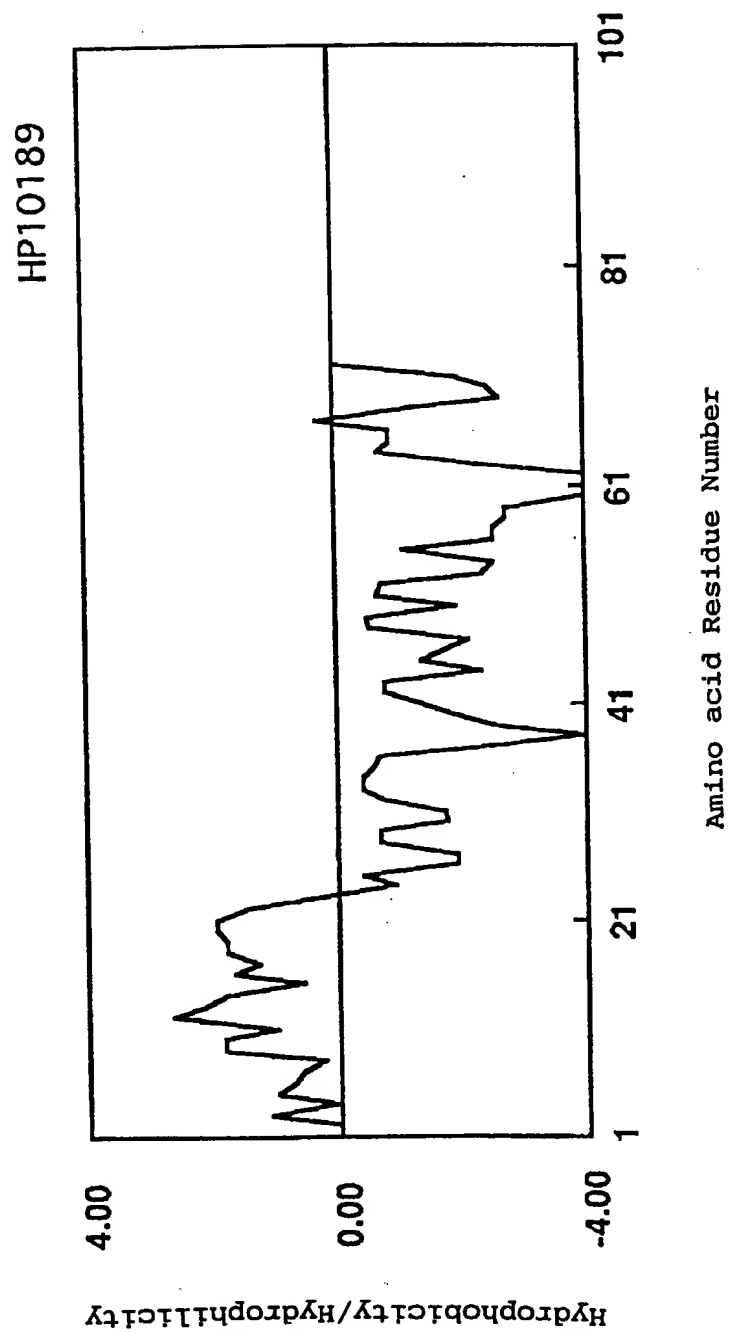
7/11

Fig. 7



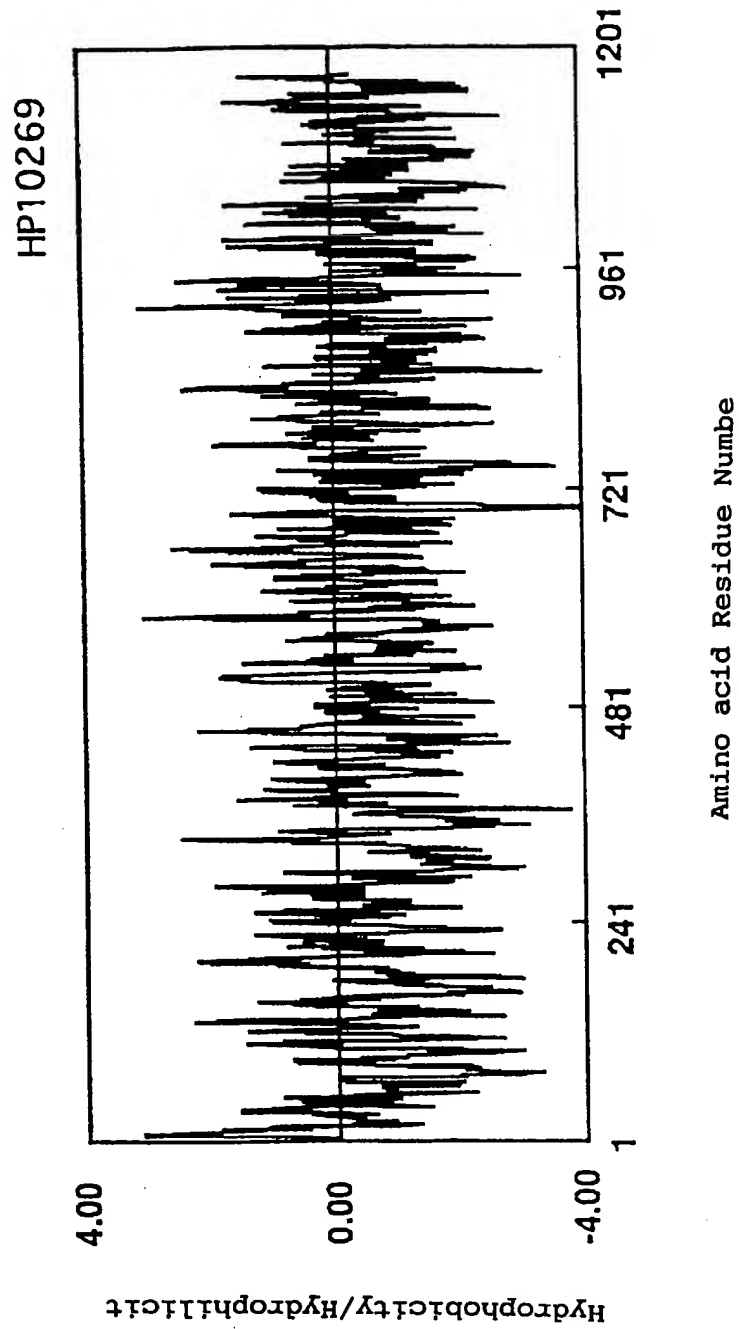
8/11

Fig. 8



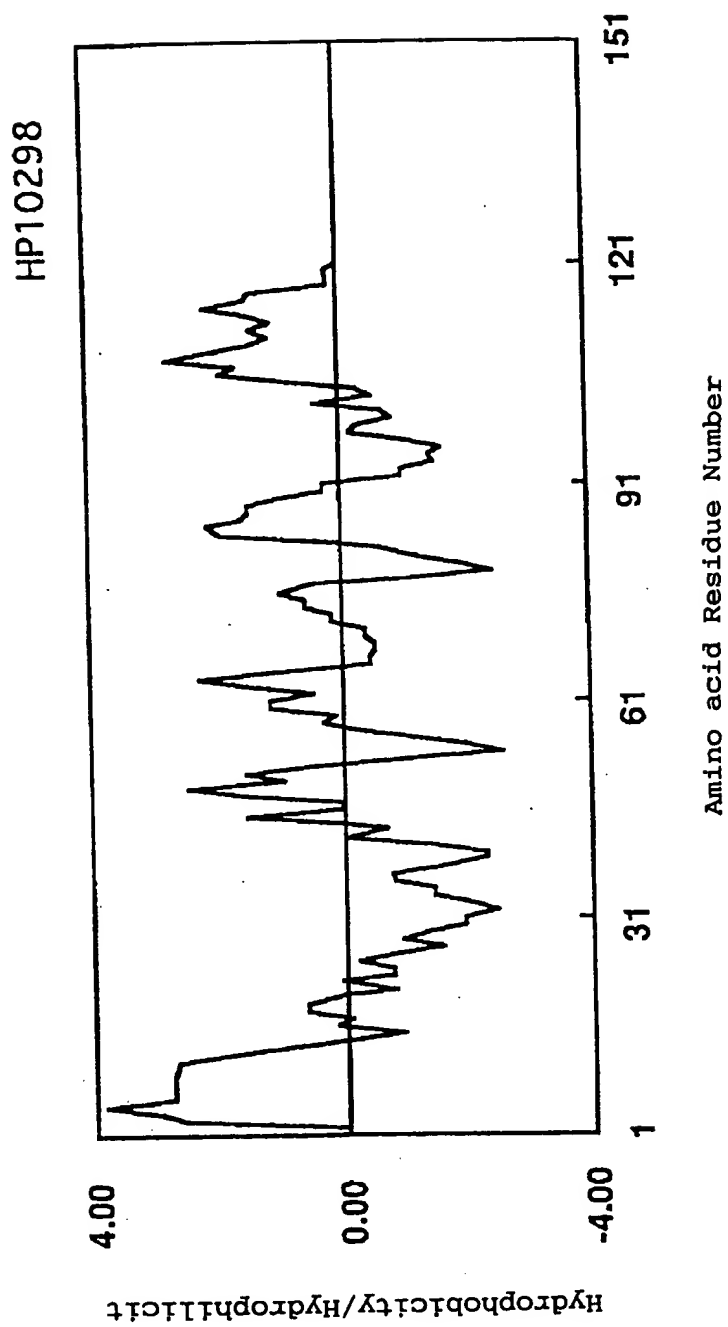
9/11

Fig. 9



10/11

Fig. 10



11/11

fig.11

